日　本　国　特　許　庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出願年月日
Date of Application: ２００１年１１月　９日

出願番号
Application Number: 特願２００１－３４４０１０

[ST.10/C]: [ＪＰ２００１－３４４０１０]

出願人
Applicant(s): 株式会社日立製作所

USSN 10/082,303
Mattingly Stanger Malur
703 684-1120
Dkt ASA-1068

CERTIFIED COPY OF
PRIORITY DOCUMENT

２００２年　２月１５日

特許庁長官
Commissioner,
Japan Patent Office

及川耕造

出証番号　出証特２００２－３００７６２７

【書類名】　特許願
【整理番号】　K01010341A
【あて先】　特許庁長官
【国際特許分類】　G06F　3/06
【発明者】
【住所又は居所】　神奈川県川崎市麻生区王禅寺１０９９番地　株式会社日立製作所　システム開発研究所内
【氏名】　本田　聖志
【発明者】
【住所又は居所】　神奈川県小田原市国府津２８８０番地　株式会社日立製作所　ストレージ事業部内
【氏名】　高安　厚志
【発明者】
【住所又は居所】　神奈川県小田原市国府津２８８０番地　株式会社日立製作所　ストレージ事業部内
【氏名】　齊木　栄作
【特許出願人】
【識別番号】　000005108
【氏名又は名称】　株式会社日立製作所
【代理人】
【識別番号】　100075096
【弁理士】
【氏名又は名称】　作田　康夫
【先の出願に基づく優先権主張】
【出願番号】　特願2001-153345
【出願日】　平成13年 5月23日
【手数料の表示】
【予納台帳番号】　013088
【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9902691

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　記憶装置システム

【特許請求の範囲】

【請求項１】

上位装置と接続された複数の記憶装置からなる記憶装置システムにおいて、

前記複数の記憶装置に含まれる第１の記憶装置が、

前記上位装置における前記記憶システムに対する情報処理についての要求を受信する手段と、

受信された前記要求を、前記複数の記憶装置に含まれる第２の記憶装置に転送可能な手段と、

受信された前記要求が、当該第１の記憶装置が実行すべき要求である場合に、受信された前記要求が示す情報処理を実行する手段とを有することを特徴とする記憶装置システム。

【請求項２】

請求項１に記載の記憶装置システムにおいて、

前記情報処理を実行する手段は、当該第１の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を実行すべきと判断された場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項３】

請求項２に記載の記憶装置システムにおいて、

前記要求には、前記情報処理を実行すべき記憶装置を示す第１の識別情報が含まれ、前記強調制御情報には、当該第１の記憶装置を識別する第２の識別情報が含まれ、

前記情報処理を実行する手段は、前記第１の識別情報と前記第２の識別情報が対応する場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項４】

請求項１乃至３のいずれかに記載の記憶装置システムにおいて、

前記第２の記憶装置が、

前記転送される要求を受信する手段と、

転送された前記要求が、当該第２の記憶装置で実行すべき要求である場合に、転送された前記要求が示す情報処理を実行する手段とを有することを特徴とする記憶装置システム。

【請求項５】

請求項４に記載の記憶装置システムにおいて、

前記第１の記憶装置では、転送可能な手段が、転送する要求に当該第１の記憶装置を示す情報を付加し、

前記第２の記憶装置は、付加された前記第１の記憶装置を示す情報に基づいて、転送された前記要求の再度の転送を抑止する手段を、さらに有することを特徴とする記憶装置システム。

【請求項６】

請求項１に記載の記憶装置システムにおいて、

前記転送可能な手段は、当該第１の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を実行すべきでないと判断された場合に、受信された前記要求を、前記第２の記憶装置に転送することを特徴とする記憶装置システム。

【請求項７】

請求項１または６のいずれかに記載の記憶装置システムにおいて、

前記転送可能な手段は、当該第１の記憶装置が実行すべき要求を示す協調制御情報と受信された前記要求に基づいて、受信された前記要求を前記第２の記憶装置が実行すべきと判断された場合に、受信された前記要求を、前記第２の記憶装置に転送することを特徴とする記憶装置システム。

【請求項８】

請求項７に記載の記憶装置システムにおいて、

前記要求には、前記情報処理を実行すべき記憶装置を示す第１の識別情報が含まれ、前記強調制御情報には、当該第１の記憶装置を識別する第２の識別情報が含まれ、

前記情報処理を実行する手段は、前記第１の識別情報と前記第２の識別情報が

対応する場合に、前記情報処理を実行することを特徴とする記憶装置システム。

【請求項９】

請求項１および６乃至８のいずれかに記載の記憶装置システムにおいて、

前記第２の記憶装置が、

前記転送される要求を受信する手段と、

転送された前記要求が、当該第２の記憶装置で実行すべき要求である場合に、転送された前記要求が示す情報処理を実行する手段とを有することを特徴とする記憶装置システム。

【請求項１０】

請求項９に記載の記憶装置システムにおいて、

前記第１の記憶装置では、転送可能な手段が、転送する要求に当該第１の記憶装置を示す情報を付加し、

前記第２の記憶装置は、付加された前記第１の記憶装置を示す情報に基づいて、転送された前記要求の再度の転送を抑止する手段を、さらに有することを特徴とする記憶装置システム。

【請求項１１】

請求項１乃至１０のいずれかに記載の記憶装置システムにおいて、

前記要求は、前記複数の記憶装置のいずれかに記憶された情報に対する読み出し要求および前記複数の記憶装置のいずれかへの情報の書き込み要求のうち少なくとも一方が含まれることを特徴とする記憶装置システム。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、データを二重化し保持する記憶装置システムに係る。特に、当該記憶装置システムを構成する複数の記憶装置間での協調制御に関する。

【０００２】

【従来の技術】

“A Case for Redundant Array of Inexpensive Disks (RAID)” (Proceedings of ACM SIGMOD, 1988)には、複数のディスク装置から構成されるディスクアレ

イ装置における、冗長データの生成、及び格納方法による分類が規定されている。

【０００３】

上記の技術によれば、二台のディスク装置に対して二重化したデータを格納することで、冗長構成を取る一台のディスク装置において障害が発生し、当該ディスク装置の格納データの読出し、或いは、当該ディスク装置に対する書込みが不可能になった場合でも、他のディスク装置を用いてデータの読出し、書込みを可能とするディスクアレイ装置は、RAID1と呼んでいる。

【０００４】

更に、上記RAID1のディスクアレイ装置において、ディスクアレイ装置内のディスク制御装置から二重化したディスク装置に対するライトデータの転送処理を軽減することで、ディスク制御装置の処理性能を向上させる技術が、特開平10-74129（第一の従来技術）に開示されている。

【０００５】

第一の従来技術によれば、データライトの際、ディスク制御装置は、第一のディスク装置にのみライトデータを転送し、該ライトデータを第一のディスク装置が、ディスク制御装置を経由せず、第二のディスク装置に転送することによって、ディスク制御装置から第二のディスク装置に対するライトデータの転送処理を不要とすることが可能となる。

【０００６】

更に、二台のディスク装置間を専用の連絡線で相互に結合し、データライトの際、相互に連絡を取りながら同時にライト処理を実施する方法が、特開平7-281959（第二の従来技術）に開示されている。

【０００７】

第二の従来技術によれば、二台のディスク装置に対し同時にライトデータの転送を実施することによって、インタフェース上のライトデータの転送量は、二重化しない場合と同等で実現することが可能となる。

【０００８】

【発明が解決しようとする課題】

上述の第一の従来技術においては、少なくとも、ディスクアレイ装置を構成する個々のディスク装置を制御し、第一のディスク装置に対して特殊なライトコマンドを発行する手段を持つディスク制御装置を必要とする為、ディスクアレイ装置が高価になるという課題がある。

【０００９】

更に、上記従来技術においては、インタフェース上のライトデータの転送量を低減するものではなく、即ち、ディスク制御装置とディスク装置とを接続するインタフェース負荷を軽減するものではない為、ディスクアレイ装置内のインタフェースの競合による性能低下を改善することができないという課題がある。

【００１０】

また、第二の従来技術においては、第一の従来技術で課題とされた、ディスク制御装置を不要とし、更に、インタフェース上のライトデータの転送量増加を防止するものであるが、その実現において、ディスク装置間で連絡をとりあう為の専用の連絡線が必要となる。

【００１１】

更に、二台のディスク装置で同期して、或いは、主側装置が副側装置の動作の進行を確かめながら、ライトデータの転送処理を実現する為、ディスク装置の利用効率の低下による性能低下、或いは、一台のディスク装置のデータライト処理に対し余分に処理時間が掛かることによる性能低下という課題がある。

【００１２】

また、上述の従来技術においては、第一／第二、或いは、主側／副側の切り替え制御方法について、更に、データリード処理時のディスク装置選択制御方法について、充分考慮されているとは言えない。

【００１３】

本発明の目的は、複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、上位装置からのデータリード／ライト要求を、該記憶装置システムを構成する複数の記憶装置間で協調して処理する制御方式を提供することにある。このことにより、上述の課題を解決する記憶装置システムを提供することである。

【００１４】

また、本発明は、ＲＡＩＤ１以外のＲＡＩＤシステムにも適用可能である。

【００１５】

【課題を解決するための手段】

上記目的を達成するために、複数の記憶装置から構成されるRAID1を含む記憶装置システムにおいて、以下の構成をとる。

【００１６】

本発明では、記憶装置システムに含まれる記憶装置が、上位装置における記憶装置システムに対する情報処理についての要求を受信する手段と、受信された要求を、複数の記憶装置に含まれる第２の記憶装置に転送可能な手段と、受信された要求が、当該第１の記憶装置が実行すべき要求である場合に、受信された要求が示す情報処理を実行する手段とを有するものである。

【００１７】

また、記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対を成す記憶装置との間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備する構成であってもよい。

【００１８】

【発明の実施の形態】

以下、本発明に係る第１の実施の形態を、図を用いて説明する。

図１は、二台の記憶装置１-a/bから構成される記憶装置システム４と、複数の上位装置２とが、任意のインタフェース（図中、SAN：Storage Area Network）３で接続される、情報処理システムの一構成例を示す図である。

【００１９】

同図において、記憶装置１としてディスク装置を例に構成を示しており、該ディスク装置１-a/bは、夫々前記上位装置２がアクセスするデータを保持する記憶媒体であるディスク部１１と、該ディスク部１１と前記上位装置２との間で転送されるデータを一時保持するバッファ部１２と、前記上位装置２との間のインタフェースプロトコル制御を実行するインタフェース制御部１３と、前記ディスク

部１１に対する記録再生処理を実行するディスク制御部１４と、前記バッファ部１２に対するアクセスを制御するバッファ制御部１５と、上記各部位を統括制御するメイン制御部１０とから構成されている。

【００２０】

また、同図において、前記ディスク装置１-a/bが二重化されたデータを保持することによって、前記RAID1の記憶装置システムを構成するものである。

【００２１】

尚、同図において、該記憶装置システムを二台のディスク装置から構成する場合を例としているが、本発明の記憶装置システムの構成は、これに限るものではなく、３台以上の記憶装置が含まれてもよい。

【００２２】

以下、本発明に係る第一の実施形態を、図２、３、４、５、６、７、８を用いて説明する。本実施形態は、各記憶装置が要求を有し（もしくは受信し）、各記憶装置が当該記憶装置でその要求に対応する情報処理を実行すべきか判断する。判断には、協調制御情報２８を用いる。

【００２３】

本実施形態では、上位装置２から記憶装置システム４に対するアクセス要求を、該記憶装置システム４を構成する複数の記憶装置１で共有し、更に、個々の記憶装置において、前記アクセス要求を自身が処理すべきか否かを判別することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現することを可能にする。

【００２４】

図２は、前記記憶装置システム４を構成するディスク装置１の一構成例を示す図である。

同図において、前記インタフェース制御部１３は、前記インタフェース３を介して情報の受信を実施する受信部２０と、同様に情報の送信を実施する送信部２１と、受信したフレーム等の情報に対するエラー検出、或いは、該フレーム情報の少なくとも一部を前記バッファ部１２に格納する際、格納先の制御等を実施す

る受信フレーム処理部２２と、フレーム等の情報送信の際、該フレーム情報を構成するヘッダ情報等の付加情報の生成を実施する送信フレーム生成部２３とを具備する。

【００２５】

更に、前記受信部２０で受信したフレーム等の情報を後段の装置に対し再送するか否かを判別し、再送制御信号を生成する再送判定部２４と、上記再送制御信号に基づき、前記送信フレーム生成部２３からの情報と、前記受信部２０で受信した情報との一方を選択し、前記送信部２１に出力する出力選択部２５とを具備する。

【００２６】

また、前記バッファ部１２は、上位装置２から受信したアクセス要求（コマンド）を保持する受信コマンド格納部２６と、上位装置２との間で送受信されるデータを保持する送受信データ格納部２７とを具備する。

【００２７】

また、前記メイン制御部１０では、上位装置２から記憶装置システム４に対するアクセス要求について、該アクセス要求を後段の装置に再送するか否か、更には、該アクセス要求を自身が処理すべきか否かを判別する為の情報等、対を成すディスク装置との間での協調制御情報２８を保持している。なお、協調制御情報２８の詳細に関しては、図１６を用いて後述する。

【００２８】

更にまた、前記再送判定部２４では、上位装置２から記憶装置システム４に対するフレーム等の情報の再送制御として、アクセス要求については、前記協調制御情報２８に基づく再送制御を、ライトデータについては、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を実施するものである。

【００２９】

以下、上位装置から記憶装置システムに対するアクセス要求について、アクセス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを個々のディスク装置で判別する場合のリード／ライトデータ転送処理を例に動作を説明す

る。

【００３０】

図３は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのリードデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【００３１】

(1) アクセス要求（コマンド）受信及び再送処理

上位装置２から記憶装置システム４に対して発行されたリードデータ要求は、フレーム情報として、先ず、ディスク装置１-aの前記受信部２０を介して受信される。ディスク装置１-aでは、前記リードデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。

【００３２】

また、ディスク装置１-aの前記再送判定部２４は、前記協調制御情報２８に基づき、上位装置２から記憶装置システム４に対するアクセス要求を後段のディスク装置１-bに対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部２５に対する再送制御（指示）信号を生成する。結果、前記出力選択部２５は、上記再送制御（指示）信号に基づき、前記上位装置からのリードデータ要求を後段のディスク装置１-bに対し再送する。

【００３３】

ディスク装置１-bでは、前記ディスク装置１-aによって再送された前記リードデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。また、ディスク装置１-bの前記再送判定部２４は、前記協調制御情報２８に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【００３４】

(2)アクセス要求（コマンド）解釈及びアクセス要求（コマンド）実行

ディスク装置１-a/bは、前記受信コマンド格納部２６に格納したアクセス要求

を解釈し、該アクセス要求がリードデータ要求であることを検出する。

【００３５】

更に、該リードデータ要求を構成するリード要求領域情報と、前記協調制御情報２８とに基づき、前記該リードデータ要求を自身が処理すべきか否かを判別する。

【００３６】

自身が処理すると判断したディスク装置では、前記ディスク制御部１４を介してディスク部１１から前記バッファ部１２の前記送受信データ格納部２７に対するディスクリード処理を開始し、更に、前記インタフェース制御部１３を介して上位装置２に対するリードデータ送信処理を実施する。更に、リードデータ送信後、前記リードデータ要求に対するステータス情報を生成送信し、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【００３７】

一方、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、リードデータ要求に掛かる処理を完了する。

【００３８】

なお、本実施形態では、アクセス要求が、ディスク装置1-aを介してディスク装置1-bに転送されるが、ディスク装置それぞれに、上位装置２がアクセス要求を送信してもよい。

【００３９】

図４は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２-a/bからのリードデータ要求が発行された場合、特に、上位装置２-aからのリードデータ要求がディスク装置１-aで、上位装置２-bからのリードデータ要求がディスク装置１-bで処理される場合の、リードデータ要求、及び、リードデータの転送処理等の流れを示す図である。

【００４０】

上述の説明の通り、上位装置２-a/bからのリードデータ要求は、少なくともディスク装置１-aのインタフェース制御部１３を介して、ディスク装置１-a/bのバッファ部１２に格納される。

【００４１】

ディスク装置１-a/bでは、受信したアクセス要求について、自身が処理すべきか否かを判別し、自身が処理すると判断したリードデータ要求に対して、ディスク部１１からのディスクリード処理、及び、上位装置２-a/bに対するリードデータ送信処理、更に、ステータス生成送信処理を実施する。ここで、ディスク装置１-aが送信するリードデータは、上位装置２-aに対するものであることから、従来通りディスク装置1-bのインタフェース制御部１３を介して上位装置２-aに転送される。

【００４２】

図５は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【００４３】

(1)アクセス要求（コマンド）受信及び再送処理

対象となるアクセス要求が、リードデータ要求でなくライトデータ要求であること以外は、図３を用いて上述した内容と同じである。上位装置２から記憶装置システム４に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置１-aの前記受信部２０を介して受信される。

【００４４】

ディスク装置１-aでは、前記ライトデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。

【００４５】

また、ディスク装置１-aの前記再送判定部２４は、前記協調制御情報２８に基づいて、上位装置２から記憶装置システム４に対するアクセス要求を後段のディスク装置１-bに対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部２５に対する再送制御（指示）信号を生成する。結果、前記出力選択部２５は、該再送制御（指示）信号に基づき、前記上位装置からのライトデータ要求を後段のディスク装置１-bに対し再送する。

【００４６】

ディスク装置１-bでは、前記ディスク装置１-aによって再送された前記ライトデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。また、ディスク装置１-bの前記再送判定部２４は、前記協調制御情報２８に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【００４７】

(2) アクセス要求（コマンド）解釈及びアクセス要求（コマンド）実行

対象となるアクセス要求が、リードデータ要求でなくライトデータ要求であること以外は、図３を用いて上述した内容と同じである。ディスク装置１-a/bは、前記受信コマンド格納部２６に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出する。更に、該ライトデータ要求を構成するライト要求領域情報と、前記協調制御情報２８とに基づき、前記該ライトデータ要求を自身が処理すべきか否かを判別する。

【００４８】

自身が処理すると判断したディスク装置では、前記バッファ部１２の送受信データ格納部２７を確保し、上位装置に対しライトデータの転送開始要求を送信する。

【００４９】

更に、上記ライトデータの転送開始要求を受信した上位装置２から送信されるライトデータを、予め確保した送受信データ格納部２７に格納する。更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部１１に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報を生成送信し、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【００５０】

一方、自身が処理しないと判断したディスク装置では、前記アクセス要求を構成するライト要求領域情報に基づき、後述の未更新領域管理情報を更新した後、前記アクセス要求を削除することによって、ライトデータ要求処理を完了する。

【００５１】

図６は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２-a/bからのライトデータ要求が発行された場合、特に、上位装置２-aからのライトデータ要求がディスク装置１-aで、上位装置２-bからのライトデータ要求がディスク装置１-bで処理される場合の、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れを示す図である。

【００５２】

上述の説明の通り、上位装置２-a/bからのライトデータ要求は、少なくともディスク装置１-aのインタフェース制御部１３を介して、ディスク装置１-a/bのバッファ部１２に格納される。

【００５３】

ディスク装置１-a/bでは、受信したアクセス要求について、自身が処理すべきか否かを判別し、　自身が処理すると判断したライトデータ要求の要求元上位装置に対して、ライトデータの転送開始要求を送信し、これを契機に上位装置から送信が開始されるライトデータを、送受信データ格納部２７に格納する。

更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部１１に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報の生成送信処理を実施する。

【００５４】

ここで、ディスク装置１-aから上位装置２-aに対して送信されるライトデータの転送開始要求、及び、ステータス情報は、前述のリードデータの転送処理と同様に、従来通りディスク装置1-bのインタフェース制御部１３を介して上位装置２-aに転送される。

【００５５】

また、上位装置２-a/bから記憶装置システム４に対して送信されるライトデータについては、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を、前記ディスク装置1-a/bのインタフェース制御部１３の再送判定部２４で実施する。即ち、上位装置２-aからのライトデータについては、前記ディスク装置1-aのインタフェース制御部１３の再送判定部２４で、該ライトデ

ータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段に転送されること無くディスク装置１-aのバッファ部１２に格納される。

【００５６】

また、上位装置２-bからのライトデータについては、前記ディスク装置1-aのインタフェース制御部１３の再送判定部２４で、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段への転送が実施され、更に、前記ディスク装置1-b′のインタフェース制御部１３の再送判定部２４で、該ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）判定制御の結果、後段に転送されること無くディスク装置１-bのバッファ部１２に格納される。

【００５７】

上述の通り、本実施形態では、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する個々の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を、前記複数の記憶装置で共有する手段と、共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべきと判断した記憶装置に転送する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な記憶装置システムが実現可能となる。

【００５８】

尚、本実施形態では、上位装置２からのライトデータ要求によって、一時的にディスク装置１-a/bに格納されたデータの不一致が生じてしまう。そこで、以下に本実施形態におけるライトデータの更新処理例を説明する。

【００５９】

図７は、前述の上位装置からのライトデータ要求を自身が処理しないと判断したディスク装置において、更新処理が実施される未更新領域管理情報３０の一構成例を示す図である。

【００６０】

未更新領域管理情報３０としては、少なくとも未更新領域に関する情報を持つ

ものであり、同図においては、上位装置から記憶装置システムに対するアクセス要求が、アクセス開始アドレス情報と、アクセスサイズ情報とから構成される場合を例に、未更新開始アドレス情報と、未更新サイズ情報とを持ち、更に、未更新領域に関する更新時刻情報等の更新情報を含む構成としている。

【００６１】

上位装置からのライトデータ要求を受信し、自身が処理しないと判断したディスク装置では、ライトデータ要求を構成するライト要求領域情報に基づき、未更新領域管理情報３０を更新することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を可能としている。

【００６２】

図８は、ディスク装置１-bの未更新領域管理情報３０に基づき、不一致の発生した未更新領域の更新処理を実施する場合の、更新要求、及び、更新データの転送処理等の流れを示す図である。

同図において、ディスク装置１-bは、自身の未更新領域管理情報３０に基づき、対を成すディスク装置１-aに対し、更新要求を発行する。該更新要求を受領したディスク装置１-aでは、必要ならばディスク部１１からのディスクリードを実施し、要求対象のデータをディスク装置１-bに対し送信する。ディスク装置１-bでは、上記ディスク装置１-aから受信したデータをディスク部１１に格納し、当該処理に掛かる未更新領域情報を未更新領域管理情報３０から削除することによって、未更新領域の更新処理を完了する。

【００６３】

尚、上述の実施形態において、上位装置からのライトデータ要求を受信し、自身が処理しないと判断したディスク装置が、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実施するものとしているが、これに限るものではなく、例えば、自身が処理すると判断したディスク装置が、ライトデータ要求を構成するライト要求領域情報に基づき、未更新領域管理情報３０を更新することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実施することも可能である。

【００６４】

更に、未更新領域の更新処理を開始する契機は任意である。例えば、未更新領域管理情報３０に登録されたエントリー数、或いは、総未更新領域サイズ等に基づき開始することが可能である。

【００６５】

以下、本発明に係る第二の実施形態を、図９、１０、１１、１２、１３を用いて説明する。本実施形態では、受信したアクセス要求を、まず受信したディスク装置で処理すべきか否か判別する。判別の結果、当該ディスク装置で処理すべきものでなければ、他のディスク装置にアクセス要求を転送する、ことに特徴がある。つまり、本実施形態では、上位装置から記憶装置システムに対するアクセス要求を、記憶装置システムを構成する一台の記憶装置で受領し、更に、当該記憶装置において、アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現可能にするものである。

【００６６】

図９は、前記記憶装置システム４を構成するディスク装置１の一構成例を示す図である。同図において、上位装置２から受領したアクセス要求を、対を成すディスク装置に転送するか否かを判定制御するコマンド転送制御部２９を新たに具備したことを除いては、図２と同一の構成となっている。

【００６７】

また、前記協調制御情報２８としては、上記上位装置２から受領したアクセス要求を、対を成すディスク装置に転送するか否かの判定制御を実施する為の情報等を保持している。更に、前記再送判定部２４では、少なくとも、上位装置２から記憶装置システム４に対するフレーム等の情報の再送制御として、ライトデータのフレームを構成するヘッダ情報等に基づく再送（転送）制御を実施するものである。

【００６８】

以下、上位装置から記憶装置システムに対するアクセス要求について、アクセ

ス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを個々のディスク装置で判別する場合のリード／ライトデータ転送処理を例に動作を説明する。

【００６９】

図１０は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのリードデータ要求が発行された場合の、各装置における処理の流れを示す図である。尚、以下の説明では、ディスク装置１-aが上位装置２からのアクセス要求を受領し、該アクセス要求を処理すべき装置の判別、更に、必要な場合、前記アクセス要求をディスク装置１-bに転送に対して転送制御を実施するものとしている。

【００７０】

(1)アクセス要求（コマンド）受信

上位装置２から記憶装置システム４に対して発行されたリードデータ要求は、フレーム情報として、先ず、ディスク装置１-aの前記受信部２０、及び受信フレーム処理部２２を介して、バッファ部１２の受信コマンド格納部２６に格納される。

【００７１】

(2)アクセス要求（コマンド）解釈及びアクセス要求（コマンド）転送処理

ディスク装置１-aは、上記受信コマンド格納部２６に格納したアクセス要求を解釈し、該アクセス要求がリードデータ要求であることを検出する。更に、該リードデータ要求を構成するリード要求領域情報と、前記協調制御情報２８とに基づき、該リードデータ要求を自身が処理すべきか否かを判別する。

【００７２】

上記判定で、自身が処理すると判断した場合、アクセス要求はディスク装置１-aで処理される。また、自身で処理しないと判断した場合、ディスク装置１-aは、前記上位装置２からのリードデータ要求をディスク装置１-bに対し転送する。

【００７３】

ディスク装置１-bでは、ディスク装置１-aから転送されたリードデータ要求を、受信部２０、及び受信フレーム処理部２２を介して、バッファ部１２の受信コ

マンド格納部２６に格納する。更に、ディスク装置１-bは、受信コマンド格納部２６に格納したアクセス要求を解釈し、アクセス要求がリードデータ要求であることを検出する。

【００７４】

(3)アクセス要求（コマンド）実行

自身が処理すると判断したディスク装置１-a、或いは、ディスク装置１-aからアクセス要求を転送されたディスク装置１-bでは、リードデータ要求に基づき、前記ディスク制御部１４を介してディスク部１１からバッファ部１２の送受信データ格納部２７に対するディスクリード処理を開始する。更に、前記インタフェース制御部１３を介して上位装置２に対するリードデータ送信処理を実施する。更に、リードデータ送信後、リードデータ要求に対するステータス情報を生成送信し、アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。尚、ディスク装置１-aにおいて自身が処理しないと判断した場合、リードデータ要求を削除することによって、リードデータ要求に掛かる処理を完了する。

【００７５】

図１１は、ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２-a/bからのリードデータ要求が発行された場合、特に、上位装置２-aからのリードデータ要求がディスク装置１-aで、上位装置２-bからのリードデータ要求がディスク装置１-bで処理される場合の、リードデータ要求、及び、リードデータの転送処理等の流れを示す図である。

【００７６】

上述の説明の通り、上位装置２-aからのリードデータ要求は、ディスク装置１-aによって受領され、ディスク装置１-bに転送されることは無い。また、上位装置２-bからのリードデータ要求は、上述のディスク装置１-aの転送制御処理によって、ディスク装置１-bが受領する。

【００７７】

ディスク装置１-a/bでは、受領したリードデータ要求に対して、ディスク部１１からのディスクリード処理、及び、上位装置２-a/bに対するリードデータ送信

処理、更に、ステータス生成送信処理を実施する。

【００７８】

ここで、ディスク装置１-aが送信するリードデータ等は、上位装置２-aに対するものであることから、従来通りディスク装置1-bのインタフェース制御部１３を介して上位装置２-aに転送される。

【００７９】

図１２は、ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【００８０】

(1) アクセス要求（コマンド）受信

上位装置２から記憶装置システム４に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置１-aの前記受信部２０、前記受信フレーム処理部２２を介して、バッファ部１２の前記受信コマンド格納部２６に格納される。

【００８１】

(2) アクセス要求（コマンド）解釈及びアクセス要求（コマンド）転送処理

ディスク装置１-aは、受信コマンド格納部２６に格納したアクセス要求を解釈し、アクセス要求がライトデータ要求であることを検出する。更に、該ライトデータ要求を構成するライト要求領域情報と、前記協調制御情報２８とに基づき、該ライトデータ要求を自身が処理すべきか否かを判別する。

【００８２】

この判定で、自身が処理すると判断した場合、前記アクセス要求はディスク装置１-aで処理される。また、自身で処理しないと判断した場合、ディスク装置１-aは、上位装置２からのライトデータ要求をディスク装置１-bに転送する。

【００８３】

ディスク装置１-bでは、ディスク装置１-aによって転送された前記ライトデータ要求を、前記受信部２０、及び前記受信フレーム処理部２２を介して、バッファ部１２の前記受信コマンド格納部２６に格納する。更に、ディスク装置１-bは

、受信コマンド格納部２６に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出する。

【００８４】

(3) アクセス要求（コマンド）実行　自身が処理すると判断したディスク装置１-a、或いは、ディスク装置１-aからアクセス要求を転送されたディスク装置１-bでは、前記ライトデータ要求に基づき、バッファ部１２の送受信データ格納部２７を確保し、上位装置に対しライトデータの転送開始要求を送信する。更に、上記ライトデータの転送開始要求を受信した上位装置から送信されるライトデータを、上記予め確保した送受信データ格納部２７に格納する。

【００８５】

更に、ライトデータ受信後、或いは、受信したライトデータを前記ディスク部１１に対しライトするディスクライト処理完了後、前記ライトデータ要求に対するステータス情報を生成送信する。

【００８６】

(4) アクセス要求（コマンド）完了処理

ディスク装置１-bでは、ステータス情報を生成送信の後、アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。また、ディスク装置１-aでは、上記ステータス情報を生成送信の後、或いは、自身で処理しないと判断し、上位装置２からのライトデータ要求をディスク装置１-bに対し転送した後、ライトデータ要求を構成するライト要求領域情報に基づき、後述の未更新領域管理情報を更新する。更に、アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【００８７】

本実施形態において、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２-a/bからのライトデータ要求が発行された場合、特に、上位装置２-aからのライトデータ要求がディスク装置１-aで、上位装置２-bからのライトデータ要求がディスク装置１-bで処理される場合の、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れは、図６と同一であることから省略する。

【００８８】

但し、上述の説明の通り、上位装置２-a/bからのアクセス要求は、少なくともディスク装置１-aのバッファ部１２に格納され、ディスク装置１-aが自身で処理しないと判断した場合に、ディスク装置１-bに対して転送されるものである。

【００８９】

上述の通り、本実施形態では、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を受領する。更に、当該記憶装置において、アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送する手段を具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、最適な記憶装置の利用効率が期待される高性能な記憶装置システムが実現可能となる。

【００９０】

また、前述の実施形態と同様に、本実施形態でも上位装置２からのライトデータ要求によって、一時的にディスク装置１-a/bに格納されたデータの不一致が生じてしまう。そこで、以下に本実施形態におけるライトデータの更新処理例を説明する。

【００９１】

本実施形態では、前述の未更新領域管理情報３０対して、新たに自身の格納するデータが最新か否かの情報（以下、更新種別情報）を付加した未更新領域管理情報をディスク装置１-aで保持し、ディスク装置１-aにおいて、上位装置からのライトデータ要求を構成するライト要求領域情報に基づき、上記未更新領域管理情報を構成する未更新領域情報と、更新種別情報とを更新・管理することによって、不一致の発生した未更新領域の管理、及び、以降の未更新領域の更新処理を実現するものである。

【００９２】

図１３は、前記ディスク装置１-aの未更新領域管理情報に基づき、不一致の発生した未更新領域の更新処理を実施する場合の、更新要求、及び、更新データの

転送処理等の流れを示す図である。また、同図において、ディスク装置１-aがディスク装置１-bに格納された最新のデータを用いて、自身を更新する場合を例としている。

【００９３】

ディスク装置１-aは、少なくとも、未更新領域情報と更新種別情報とから構成される前記未更新領域管理情報に基づき、対を成すディスク装置１-bに対し、更新要求（最新データ送信要求）を発行する。該更新要求を受領したディスク装置１-bでは、必要ならば前記ディスク部１１からのディスクリードを実施し、要求対象のデータをディスク装置１-aに対し送信する。ディスク装置１-aでは、上記ディスク装置１-bから受信したデータをディスク部１１に格納し、当該処理に掛かる未更新領域情報を前記未更新領域管理情報３０から削除することによって、未更新領域の更新処理を完了する。

【００９４】

また、ディスク装置１-aが自身の格納する最新データを用いて、ディスク装置１-bを更新する場合、ディスク装置１-aは、前記未更新領域管理情報に基づき、対を成すディスク装置１-bに対し、更新要求（最新データ受信要求）を発行し、引き続き更新対象のデータを送信する。更新要求及び更新対象データを受領したディスク装置１-bでは、上記ディスク装置１-aから受信したデータをディスク部１１に格納する。更に、ディスク装置１-aは、更新対象データの送信後、当該処理に掛かる未更新領域情報を前記未更新領域管理情報３０から削除することによって、未更新領域の更新処理を完了する。

【００９５】

以上の実施形態では、上位装置から記憶装置システムに対するアクセス要求を構成するアクセス要求領域を条件として、該アクセス要求を自身が処理すべきか否かを判別する場合を例に動作を説明してきたが、これに限定するものではない。

【００９６】

例えば、記憶装置がオブジェクト（ファイル）単位で格納するデータを管理する場合、上位装置から記憶装置システムに対するアクセス要求を構成するオブジ

ェクト（ファイル）情報を条件として、該アクセス要求を自身が処理すべきか否かを判別することが可能であり、前記未更新領域管理情報として少なくとも保持する未更新領域情報も、オブジェクト（ファイル）情報を用いることで、上述の未更新領域の更新処理を実現することが可能となる。

【００９７】

また、以上の実施形態では、複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、当該記憶装置システムを構成する個々の記憶装置が、或いは、少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求について、自身が処理すべきか否かを判別する手段を具備することによって、個々の記憶装置を制御する制御装置を不要とする。また、最適な記憶装置の利用効率が期待される高性能な前記記憶装置システムを実現するものである。

【００９８】

しかし、最適な記憶装置の利用効率を実現する為、上位装置からのライトデータ要求によって、一時的にではあるが、二重化されたディスク装置のデータに不一致が生じ、上記ライトデータの更新処理を必要とする場合もある。また、上記上記ライトデータの更新処理において、前記インタフェース３を介したライトデータの転送処理を実施する為、インタフェース３の負荷を軽減するものではない。

【００９９】

そこで、上位装置２から記憶装置システム４に対するアクセス要求、特に、ライトデータ要求を、該記憶装置システム４を構成する複数の記憶装置１で共有し、更に、対を成す複数の記憶装置が同期して該ライトデータ要求を処理することによって、インタフェース負荷を軽減することによる高性能化を図った実施形態について、以下に説明する。

【０１００】

以下、本発明に係る第三の実施形態を、図１４、１５を用いて説明する。

本実施形態では、上位装置２から記憶装置システム４に対するライトデータ要求を、該記憶装置システム４を構成する複数の記憶装置１で共有し、更に、を成す複数の記憶装置でライトデータの転送処理をも同期して実施することによって

、個々の記憶装置を制御する制御装置が不要で低価格な、また、インタフェース負荷を軽減する高性能な前記記憶装置システムを実現するものである。

【０１０１】

尚、本実施形態の記憶装置システムを構成する記憶装置（ディスク装置）の構成は、再送判定制御部２４において、上位装置２から記憶装置システム４に対するフレーム等の情報の再送制御として、ライトデータについても、前記協調制御情報２８に基づく再送制御を実施することを除いて、図２と同一であることから省略する。

【０１０２】

また、上位装置から記憶装置システムに対するリードデータ要求については、本実施形態の対象外であり、上述の実施形態の何れかを採用するものとして、その説明も省略する。

【０１０３】

以下、上位装置から記憶装置システムに対するライトデータ要求について、対を成す二台の記憶装置でライトデータの転送処理を同期して実施する場合のライトデータ転送処理を例に動作を説明する。

図１４は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのライトデータ要求が発行された場合の、各装置における処理の流れを示す図である。

【０１０４】

図１５は、前記ディスク装置１-a/bから構成される記憶装置システム４に対し、上位装置２からのライトデータ要求が発行された場合の、ライトデータ要求、ライトデータの転送開始要求、及び、ライトデータの転送処理等の流れを示す図である。

【０１０５】

(1) アクセス要求（コマンド）受信及び再送処理

上位装置２から記憶装置システム４に対して発行されたライトデータ要求は、フレーム情報として、先ず、ディスク装置１-aの前記受信部２０を介して受信される。

ディスク装置１-aでは、前記ライトデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。

【０１０６】

また、ディスク装置１-aの前記再送判定部２４は、前記協調制御情報２８に基づき、上位装置から記憶装置システム４に対するアクセス要求を後段のディスク装置１-bに対して再送するように設定しておくことによって、受信したフレーム情報がアクセス要求であることを検出し、前記出力選択部２５に対する再送制御（指示）信号を生成する。結果、前記出力選択部２５は、該再送制御（指示）信号に基づき、前記上位装置からのライトデータ要求を後段のディスク装置１-bに対し再送する。

【０１０７】

ディスク装置１-bでは、前記ディスク装置１-aによって再送された前記ライトデータ要求を、前記受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。また、ディスク装置１-bの前記再送判定部２４は、前記協調制御情報２８に基づき、アクセス要求を後段の装置に対し再送しないよう設定されている。

【０１０８】

(2) アクセス要求（コマンド）解釈及びアクセス要求（コマンド）実行

ディスク装置１-a/bは、前記受信コマンド格納部２６に格納したアクセス要求を解釈し、該アクセス要求がライトデータ要求であることを検出し、バッファ部１２の送受信データ格納部２７を確保する。ディスク装置１-aでは、自身の送受信データ格納部２７を確保したことを契機として、ディスク装置１-bに対しライトデータの転送開始要求を送信する。一方、ディスク装置１-bでは、自身の送受信データ格納部２７を確保したことと、ディスク装置１-aからのライトデータの転送開始要求受信を契機として、上位装置２に対しライトデータの転送開始要求を送信する。

【０１０９】

上位装置２では、上記ライトデータの転送開始要求を受信したことを契機として、ライトデータの送信を開始する。

上位装置２から送信されたライトデータは、先ず、ディスク装置１-aの前記受信部２０を介して受信される。ディスク装置１-aでは、上記ライトデータを、前記受信フレーム処理部２２を介して、バッファ部１２の前記送受信データ格納部２７に格納する。

【０１１０】

また、ディスク装置１-aの再送判定部２４は、協調制御情報２８に基づき、上位装置２から記憶装置システム４に対するライトデータを後段のディスク装置１-bに対して再送するように設定しておくことによって、受信したフレーム情報がライトデータであることを検出し、出力選択部２５に対する再送制御（指示）信号を生成する。出力選択部２５は、再送制御（指示）信号に基づき、上位装置からのライトデータを後段のディスク装置１-bに対し再送する。

【０１１１】

ディスク装置１-bでは、ディスク装置１-aによって再送されたライトデータを、受信フレーム処理部２２を介して、バッファ部１２の前記送受信データ格納部２７に格納する。また、ディスク装置１-bの再送判定部２４は、協調制御情報２８に基づき、ライトデータを後段の装置に対し再送しないよう設定されている。

【０１１２】

ディスク装置１-aでは、ライトデータ受信後、或いは、受信したライトデータをディスク部１１に対しライトするディスクライト処理完了後、ディスク装置１-bに対し、前記ライトデータ要求に対するステータス情報を生成送信する。更に、当該ライトデータ要求を削除することによって、一連のライトデータ要求処理を完了する。

【０１１３】

ディスク装置１-bでは、ライトデータ受信後、或いは、受信したライトデータをディスク部１１に対しライトするディスクライト処理完了後、ディスク装置１からのステータス情報を受信後、上位装置２に対して当該ライトデータ要求に対するステータス情報を生成送信し、更に、当該ライトデータ要求を削除することによって、一連のライトデータ要求処理を完了する。

【０１１４】

上述の通り、本実施形態では、複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、該記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対を成す複数の記憶装置間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で低価格な、また、記憶装置間でのデータ更新処理を不要とし、インタフェース負荷を軽減する高性能な前記記憶装置システムが実現可能となる。

【０１１５】

以上の実施形態では、本発明を適用した複数の記憶装置のみから構成される、RAID1の記憶装置システムについて述べてきたが、これに限定されるものではない。例えば、従来のディスクアレイ装置、即ち、複数のディスク装置と、これらのディスク装置を制御する制御装置とから構成されるディスクアレイ装置においても、本発明を適用したディスク装置を採用することによって、上記制御装置の構成が容易となり、結果、低価格化を実現することが可能となる。

【０１１６】

更に、前述の上位装置からのアクセス要求を自身が処理すべきか否かの判別処理の一例を、以下に説明する。

【０１１７】

図１６は、前記協調制御情報２８を構成する情報の一例を示す図である。

同図において、協調制御情報２８を構成する情報として、前記記憶装置システム４の装置識別情報と、個々の記憶装置自身に固有の装置識別情報と、自身と協調制御を実施する記憶装置に固有の装置識別情報と、協調処理モード情報と、更に、協調処理対象となる領域管理情報として、協調領域サイズ情報と、アクティブ領域開始アドレス情報と、アクティブ領域サイズ情報とを保持するものである。

【０１１８】

また、上記各情報について、記憶装置１-a/bにおける設定例を示しており、前

記記憶装置システム４の装置識別情報は、記憶装置１-a/bともにID_0、個々の記憶装置自身に固有の装置識別情報は、記憶装置１-aにID_1、記憶装置１-bにID_2、自身と協調制御を実施する記憶装置に固有の装置識別情報は、記憶装置１-aにID_2、記憶装置１-bにID_1を設定し、協調処理モード情報は、記憶装置１-aにコマンド再送、記憶装置１-bにコマンド受領を設定し、更に、協調領域サイズ情報はＮとして、アクティブ領域開始アドレス情報は、記憶装置１-aに０、記憶装置１-bにＮ／２、アクティブ領域サイズ情報は、記憶装置１-a/bともにＮ／２を設定したものである。

【０１１９】

尚、上記記憶装置システム４の装置識別情報と、個々の記憶装置自身に固有の装置識別情報と、自身と協調制御を実施する記憶装置に固有の装置識別情報と、協調処理モード情報とは、後述のフレームデータ再送処理判定で用いられる協調制御情報である。

【０１２０】

図１７は、前記上位装置からのアクセス要求情報として、FCP (Fibre Channel Protocol)で規定されるアクセス要求情報の構成例を示す図である。同図において、上記アクセス要求情報は、アクセス要求対象領域の開始アドレス情報（図中、Logical Block Address）と、アクセス要求対象領域サイズ情報（Transfer Length）を含むものである。

【０１２１】

上位装置からのアクセス要求情報を受領した記憶装置１は、自身の保持する前記協調制御情報の、アクティブ領域開始アドレス情報と、アクティブ領域サイズ情報と、上記上位装置からのアクセス要求情報の、アクセス要求対象領域の開始アドレス情報と、アクセス要求対象領域サイズ情報とに基づき、該アクセス要求を自身が処理すべきか否か、即ち、自身のアクティブ領域に対するアクセス要求か否かを判別し、自身のアクティブ領域に対するアクセス要求についてのみ、処理を実施する。

【０１２２】

例えば、上記上位装置からのアクセス要求対象領域が、Ｎ／２以下の場合、記

憶装置＃１が、また、Ｎ／２以上の場合、記憶装置＃２が、該上位装置からのアクセス要求を処理する。更に、上記上位装置からのアクセス要求対象領域が、記憶装置＃１及び記憶装置＃２のアクティブ領域に係る場合、記憶装置＃１及び記憶装置＃２が、夫々自身のアクティブ領域に対するアクセス要求を処理する。

【０１２３】

また、上述の実施形態では、協調処理対象領域の管理情報として、アドレス情報と、サイズ情報とを保持し、上位装置からのアクセス要求情報を構成するアドレス情報と、サイズ情報とに基づき、該アクセス要求を自身が処理すべきか否かを判別する場合を例に説明したが、これに限るものではなく、例えば、図１７に示す論理ユニット番号（図中、FCP_LUN or Logical Unit Number）を条件とすることも可能であり、この場合、前記記憶装置１は、前記協調制御情報２８を構成する協調処理対象領域に関する管理情報として、アクティブ論理ユニット番号情報を保持するものである。

【０１２４】

更にまた、上位装置からのアクセス要求情報が、ファイル名等のオブジェクト名を保持する場合、前記記憶装置１は、前記協調制御情報２８を構成する協調処理対象領域に関する管理情報として、アクティブオブジェクトリスト情報を保持することによって、オブジェクト単位で、上記上位装置からのアクセス要求を自身が処理すべきか否かを判別することも可能である。

【０１２５】

尚、前述の第二の実施形態においては、上位装置からのアクセス要求対象領域が、ディスク装置１-a及び前記ディスク装置１-bのアクティブ領域に係る場合、ディスク装置１-aが、上位装置からのアクセス要求情報を、ディスク装置１-bのアクティブ領域に対するアクセス要求に加工して前記ディスク装置１-bに対して送信することによって、ディスク装置１-a/bでの協調制御処理を実現することが可能である。

【０１２６】

或いは、前記ディスク装置１-bにおいて、前述の協調制御情報（アクティブ領域開始アドレス情報、アクティブ領域サイズ情報、等）を保持し、前記上位装置

からのアクセス要求情報の、アクセス要求対象領域情報とに基づき、該アクセス要求を自身が処理すべきか否か、即ち、自身のアクティブ領域に対するアクセス要求か否かを判別し、自身のアクティブ領域に対するアクセス要求について処理を実施することによって、上述の前記ディスク装置１-aが、前記上位装置からのアクセス要求情報を、ディスク装置１-bのアクティブ領域に対するアクセス要求に加工してディスク装置１-bに対して送信することなく、前記ディスク装置１-a/bでの協調制御処理を実現することが可能である。

【０１２７】

更に、前述の上位装置と記憶装置システムとの間で転送されるフレームデータ等の情報を後段の装置に対して再送するか否かの判定処理の一例を、以下に説明する。

【０１２８】

図１８は、上位装置と記憶装置システムとの間で転送される情報として、FC-PH (Fibre Channel PHYSICAL AND SIGNALING INTERFACE)で規定されるフレームヘッダ情報の構成例を示す図である。

【０１２９】

同図において、上記フレームヘッダ情報は、フレームデータの送信先／送信元の情報（D_ID：Destination_Identifier、S_ID：Source_Identifier）と、フレームデータ種別情報（R_CTL：Routing Control、TYPE：Data structure type）と、フレームデータの関連するイクスチェンジ識別情報（RX_ID：Responder Exchange_Identifier、OX_ID：Originator Exchange_Identifier）のフィールドを具備している。

【０１３０】

先ず、前記記憶装置システム４（前記ディスク装置１-a/b）から上位装置に対して送信されるフレームデータ（リードデータ、ライトデータの転送開始要求、ステータス情報、等）の場合、上記フレームヘッダ情報のフレームデータの送信先情報（D_ID）として、上位装置の識別情報が設定されており、当該フレームデータを受信したディスク装置は、再送する。

【０１３１】

そこで、以下では、上位装置から前記記憶装置システム４に対して送信される、即ち、前記フレームヘッダ情報のフレームデータの送信先情報（D_ID）として、記憶装置システム４の識別情報が設定されたフレームデータ（アクセス要求、ライトデータ）を対象に、再送判定処理の一例を説明する。

【０１３２】

(1) アクセス要求

再送判定部２４では、上位装置から記憶装置システム４に対するフレームデータ（アクセス要求）を受信した際、前記フレームヘッダ情報のフレームデータ種別情報（R_CTL、TYPE）から、当該フレームデータがアクセス要求であることを検出し、更に、前記協調制御情報２８を構成する協調処理モード設定情報（コマンド再送、且つ或いは、コマンド受領）に基づき、該受信したアクセス要求を後段の装置に対して再送するか否かを決定することで、フレームデータ（アクセス要求）の再送処理を実現する。

【０１３３】

(2) ライトデータ

再送判定部２４では、上位装置から記憶装置システム４に対するフレームデータ（ライトデータ）を受信した際、前記フレームヘッダ情報のフレームデータ種別情報（R_CTL、TYPE）から、当該フレームデータがライトデータであることを検出し、更に、前記フレームヘッダ情報のフレームデータの送信元情報（S_ID）と、イクスチェンジ識別情報（RX_ID、OX_ID）とが、後述のライト処理管理情報に登録されたものと一致するか否かを検出し、一致する場合、該ライトデータを受領し、不一致の場合、後段の装置に対して再送することで、フレームデータ（ライトデータ）の再送処理を実現する。

【０１３４】

図１９は、上記ライト処理管理情報の一構成例を示すものであり、実行中（処理未完）のライト処理の前記イクスチェンジ識別情報（RX_ID、OX_ID）と、アクセス要求発行元上位装置の識別情報（Host ID）とを関連付けたものを、ライト処理管理情報として保持する例である。同図において、例えば、同一の上位装置（Host ID（S_ID）＝ID_3）から、前記記憶装置システム４に対して送信される

ライトデータフレームは、該フレームヘッダを構成する前記イクスチェンジ識別情報（RX_ID and/or OX_ID）が、上記ライト処理管理情報に登録されたものと一致するか否かを検出することで、自身が受領すべきライトデータフレームか否かの検出が可能となる。

【０１３５】

尚、上記ライト処理管理情報の最少の構成例としては、実行中（処理未完）のライト処理の前記イクスチェンジ識別情報（OX_ID）と、アクセス要求発行元上位装置の識別情報（Host ID）とから構成することも可能であるが、上記OX_IDは上位装置側で設定する値であり、その独立性を保障されるものではない。これに対し、前記RX_IDは記憶装置側で任意の値を設定可能であることから、前記ライト処理管理情報として保持することが望ましく、更に、協調制御の対象となる記憶装置間で異なる値を取るように設定することが望ましい。

【０１３６】

次に、本発明の第四の実施の形態について説明する。

第四の実施形態は、記憶装置サブシステムを構成する個々の記憶装置が、複数の記憶装置間で協調して上位装置からのアクセス要求を処理するための制御情報（協調制御情報）と、上位装置から記憶装置サブシステムに対するアクセス要求情報とに基づき、当該アクセス要求を自身が処理すべきか否かの判別を実施するものである。

【０１３７】

第五の実施形態は、第四の実施形態の記憶装置が、更に、協調して処理する他の記憶装置の動作状態情報を追加保持することによって、前記記憶装置サブシステムを構成する記憶装置に障害が発生した場合でも、上位装置から記憶装置サブシステムに対するアクセス要求処理を実現するものである。

【０１３８】

第六の実施形態は、上述の実施形態の記憶装置が、更に、複数のインタフェース制御部を具備し、上位装置からのアクセス要求情報と、記憶装置サブシステムを構成する記憶装置間で転送されるユーザデータとの転送処理を、異なるインタフェース（伝送路）を用いて実現するものである。

【０１３９】

［第四の実施形態］　（正常動作）

以下、本発明に係る第四の実施形態を、図２０～１４を用いて説明する。

図２０は、複数(N，N>1)台の記憶装置１から構成される記憶装置サブシステム４と、上位装置２とから構成される情報処理システムの一例を示す図であり、同図において、記憶装置サブシステム４を構成する個々の記憶装置１は、前記上位装置２と、巡回型（以下、ループ）のインタフェース３を介して直接接続されている。

【０１４０】

また、同図において、記憶装置１としてディスク装置を例に構成を示しており、該ディスク装置１は、夫々前記上位装置２がアクセスするデータを保持する記憶媒体であるディスク部１１と、該ディスク部１１と前記上位装置２との間で転送されるデータを一時保持するバッファ部１２と、前記上位装置２との間のインタフェースプロトコル制御を実行するインタフェース制御部１３と、前記ディスク部１１に対する記録再生処理を実行するディスク制御部１４と、前記バッファ部１２に対するアクセスを制御するバッファ制御部１５と、上位装置からのデータライト要求に基づき受信したライトデータ（新データ）と、前記ディスク部に格納済みの更新対象データ（旧データ）とから排他的論理和演算による更新情報を生成する更新情報生成部１６と、上記各部位を統括制御するメイン制御部１０とを具備する構成である。

【０１４１】

図２１は、前記記憶装置システム４を構成するディスク装置１の一構成例を示す図である。

【０１４２】

同図において、前記インタフェース制御部１３は、前記インタフェース３を介して情報の受信を実施する受信部２０と、同様に情報の送信を実施する送信部２１と、受信したフレーム等の情報に対するエラー検出、或いは、該フレーム情報の少なくとも一部を前記バッファ部１２に格納する際、格納先の制御等を実施する受信フレーム処理部２２と、フレーム等の情報送信の際、該フレーム情報を構

成するヘッダ情報等の付加情報の生成を実施する送信フレーム生成部２３とを具備し、

更に、前記受信部２０で受信したフレーム等の情報を後段の装置に対し再送するか否か、或いは、自身が受領するか否かの制御（再送制御）を実施するフレーム受信制御部２４と、該フレーム受信制御部２４から再送された情報と、前記送信フレーム生成部２３からの情報との一方を選択し、前記送信部２１に出力するフレーム送信制御部２５とを具備する。

【０１４３】

また、前記バッファ部１２は、上位装置２から受信したアクセス要求（コマンド）を保持する受信コマンド格納部２６と、上位装置２との間で送受信されるデータを保持する送受信データ格納部２７と、前記更新情報生成部１６にて生成された更新情報を保持する更新情報格納部２８とを具備する。

【０１４４】

また、前記メイン制御部１０は、後述の協調制御情報２９と、上位装置２から記憶装置システム４に対するアクセス要求情報とに基づき、該アクセス要求に対応する自身の処理を制御する協調処理制御部３０を具備している。

【０１４５】

尚、前記フレーム受信制御部２４における再送制御は、上記協調処理制御部３０からの制御情報と、フレームを構成するヘッダ情報等に基づき実施するものである。

【０１４６】

図２２は、前記協調制御情報２９の構成、及び個々の記憶装置における設定の一例を示す図である。同図において、該協調制御情報２９は、前記記憶装置サブシステム４の装置識別情報（共有ＩＤ）と、個々の記憶装置自身に固有の装置識別情報（固有ＩＤ）と、前記巡回型インタフェース３における相対位置情報としてのポジションマップ情報と、前段から受信したアクセス要求フレームを後段の装置に対し再送するか否かモードを指定するコマンド転送モード情報とを含む。

【０１４７】

更に、冗長構成の管理情報として、ＲＡＩＤレベル情報と、冗長データ管理サ

イズと、前記記憶装置サブシステム４を構成ディスク台数構成情報であるデータディスク台数／冗長データディスク台数情報とから構成され、個々の記憶装置毎に任意の設定が実施されている。

【０１４８】

図２３は、前記巡回型インタフェース３の一例として、Fibre Channel Arbitrated Loop (FC-AL)におけるフレーム情報の構成を示す図である。

【０１４９】

同図において、フレーム情報は、フレーム情報の開始を示すSOF (Start of Frame)と、フレーム情報の種別情報、並びに、送信先／元情報等を含むヘッダ情報と、装置間で転送されるペイロード（ユーザデータ）と、上記ヘッダ情報と、ペイロードに対する冗長データであるCRC (Cyclic Redundancy Check)と、フレーム情報の終了を示すEOF (End of Frame)とから構成されている。

【０１５０】

図２４は、上記ヘッダ情報の一例を示す図であり、フレーム情報の種別情報であるR_CTL/TYPE、フレーム情報の送信先／元情報であるD_ID/S_ID等から構成されている。

【０１５１】

図２５は、上記ペイロードとして、アクセス要求情報の一例を示す図であり、アクセス要求の対象を指定する情報である論理ユニット番号(FCP_LUN)、アドレス情報(Logical Block Address)、転送長情報(Transfer Length)、アクセス要求の種別情報であるオペレーションコード(Operation Code)等から構成されている。

【０１５２】

以下、前記上位装置２から記憶装置サブシステム４に対するアクセス要求を、前記記憶装置１－１～Ｎが協調して処理する動作を説明する。

【０１５３】

(1) アクセス要求（コマンド）受信及び再送処理

図２６は、前記記憶装置１－１～Ｎから構成される記憶装置サブシステム４に対し、上位装置２からのアクセス要求が発行された場合の、各記憶装置における

処理例を示す図である。

【０１５４】

先ず、上位装置２から記憶装置サブシステム４に対して発行されたアクセス要求(Command)は、フレーム情報として、記憶装置１－１の前記受信部２０を介して受信される。

【０１５５】

記憶装置１－１では、上記アクセス要求を、前記フレーム受信制御部２４、並びに受信フレーム処理部２２を介して、前記バッファ部１２の前記受信コマンド格納部２６に格納する。

【０１５６】

また、記憶装置１－１の前記フレーム受信制御部２４は、前記協調制御情報２９の前記コマンド転送コードを上位装置２から記憶装置サブシステム４に対するアクセス要求を後段の装置に対して再送するよう（受信＆再送）に設定しておくことによって、受信したフレーム情報の前記ヘッダ情報から、該フレーム情報がアクセス要求であることを検出し、前記フレーム送信制御部２５を介して、前記上位装置からのアクセス要求を後段の装置に対し再送する（受信＆再送）。

【０１５７】

後段の記憶装置１－２～（Ｎ－１）においても、上記と同様のコマンド受信＆再送処理を実施する。

【０１５８】

最終段の記憶装置１－Ｎでは、前記協調制御情報２９の前記コマンド転送コードを上位装置２から記憶装置サブシステム４に対するアクセス要求を後段の装置に対して再送しないよう（受信のみ）に設定しておくことによって、受信したフレーム情報の前記ヘッダ情報から、該フレーム情報がアクセス要求であることを検出し、を後段の装置に対し再送することなく、前記バッファ部１２の前記受信コマンド格納部２６への格納のみ実施する。

【０１５９】

以上のアクセス要求（コマンド）受信及び再送処理によって、協調処理を行う各記憶装置１－１～Ｎで、新たなアクセス要求の転送処理を開始することなく、

上位装置２からのアクセス要求を共有することが可能となる。

【０１６０】

(2-1) リード処理（１台の記憶装置で処理実現）

図２７は、上位装置２からの前記アクセス要求が、リードデータ要求で、且つ、一台の記憶装置で処理可能な場合のデータ転送処理例を示す図であり、同図においては、記憶装置１－２がデータ送信処理(Data2送信)を実施するものとしている。

【０１６１】

各記憶装置１－１～Ｎは、前記受信コマンド格納部２６に格納したアクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を自身が処理するか否かを判別する。

【０１６２】

自身が処理すると判断した記憶装置１－２では、前記ディスク制御部１４－２を介してディスク部１１－２から前記バッファ部１２－２の前記送受信データ格納部２７－２に対するディスクリード処理を開始し、更に、前記インタフェース制御部１３－２を介して上位装置２に対するリードデータ送信処理を実施する。更に、リードデータ送信後、前記アクセス要求に対するステータス情報を生成送信し、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０１６３】

一方、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、リードデータ要求に掛かる処理を完了する。

尚、記憶装置１－２の後段に位置する記憶装置１－３～Ｎにおいては、記憶装置１－２から送信されるフレーム情報（リードデータ、ステータス）について、そのヘッダ情報から再送処理のみを実施する。

【０１６４】

(2-2) リード処理（複数台の記憶装置で処理実現）

図２８は、上位装置２からの前記アクセス要求が、リードデータ要求で、且つ、複数台の記憶装置での処理が必要な場合の、各記憶装置におけるデータ転送処

理例を示す図であり、同図においては、記憶装置１－１／２がデータ送信処理(Data1/Data2送信)を実施するものとしている。

【０１６５】

図２９は、複数台の記憶装置でのデータ転送処理が必要な場合の、データフレーム（ペイロード）情報の一構成例を示す図であり、従来、データフレーム（ペイロード）がユーザデータのみから構成されていたのに対し、上記ユーザデータを処理対象の記憶装置毎に分割したユーザデータセグメント０～ｎと、該ユーザデータセグメントのデータが有効か否かを示すフラグ(Data Valid bit)とから構成するものとしている。

【０１６６】

尚、以下の説明では、前記アクセス要求情報と前記協調制御情報２９とに基づき、前記アクセス要求を処理する記憶装置において、最前段に位置する（上位装置からの情報を最も早く受信する）記憶装置が、リードデータ／ステータス情報の送信開始を許可する場合を例に説明するが、これに限るものではなく、任意のアルゴリズムでリードデータの送信を開始する記憶装置を決定することが可能である。

【０１６７】

前述と同様に、各記憶装置１－１～Ｎは、前記受信コマンド格納部２６に格納したアクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を自身が処理するか否かを判別し、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、リードデータ要求に掛かる処理を完了する。

【０１６８】

自身が処理すると判断した記憶装置１－１／２では、前記ディスク制御部１４－１／２を介してディスク部１１－１／２から前記バッファ部１２－１／２の前記送受信データ格納部２７－１／２に対するディスクリード処理を開始する。

【０１６９】

自身がリードデータの送信の開始を許可された記憶装置１－１では、上位装置２に対するリードデータの送信処理として、例えば、前記ユーザデータセグメン

ト０に自身のリードデータを、他のユーザデータセグメント(1)にはダミーデータを格納し、更に、前記ユーザデータセグメント０に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報の送信処理を実施し、更に、該処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０１７０】

また、自身がリードデータの送信の開始を禁止された記憶装置１－２では、上位装置２に対するリードデータの送信処理として、例えば、上記記憶装置１－１から受信したリードデータフレームのユーザデータセグメント１に自身のリードデータを格納し、更に、前記ユーザデータセグメント１に対応するフラグに“有効”を設定した前記データフレーム情報の送信処理を実施し、更に、該処理に係るステータス情報と前段の装置から受信したステータス情報とに基づく新たなステータス情報を生成送信後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０１７１】

尚、上記データ送信処理を実施する最終段の記憶装置、即ち、上位装置に対する最終形のデータフレーム情報を生成する記憶装置においては、前記ユーザデータセグメントのデータが有効か否かを示すフラグ(Data Valid bit)領域を削除して、ユーザデータのみから構成されるペイロードとして上位装置に送信することが望ましい。

【０１７２】

また、上位装置２に対するステータス情報の送信処理としては、上記データ送信処理を実施する最終段の記憶装置、即ち、上位装置に対する最終形のデータフレーム情報を生成する記憶装置において、前段の装置から受信したステータス情報と自身のステータス情報とから新たなステータス情報を生成送信することによって、上位装置２に対するステータス情報の生成を実現している。

【０１７３】

(2-3) ライト処理（１台の記憶装置でデータ更新処理実現）

図３０は、上位装置２からの前記アクセス要求が、ライトデータ要求で、且つ、一台の記憶装置のデータ更新処理と一台の記憶装置の冗長データ更新処理とを実施する場合のデータ転送処理例を示す図であり、同図においては、記憶装置１－１がデータ更新処理(Data1O→Data1N)を記憶装置１－Ｎが冗長データ更新処理(DataPO→DataPN)を実施するものとしている。

【０１７４】

前述と同様に、各記憶装置１－１～Ｎは、前記受信コマンド格納部２６に格納したアクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を自身が処理するか否か、或いは、該アクセス要求に伴う冗長データ更新処理を実施するか否かを判別し、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、上記ライトデータ要求に掛かる処理を完了する。

【０１７５】

自身がデータ更新処理を実施すると判断した記憶装置１－１では、前記ディスク制御部１４－１を介して、ディスク部１１－１に格納されている更新対象のデータ（旧データ：Data1O）に対するディスクリード処理を開始する。

【０１７６】

更に、上位装置２から前記インタフェース制御部１３－１を介して受信したデータ（新データ：Data1N）と上記旧データ(Data1O)とから、例えば、排他的論理和演算結果等による更新情報(Data1N')を前記更新情報生成部１６－１で生成する。

【０１７７】

更に、前記上位装置２から受信した新データ(Data1N)を上記更新情報(Data1N')に置換え、当該更新情報を格納するセグメントに対応するフラグに“有効”を設定した前記Data Valid bit領域を付加した前記データフレーム情報の送信処理を実施する。

【０１７８】

更に、記憶装置１－１では、上記新データ(Data1N)を、前記ディスク制御部１４－１を介してディスク部１１－１に格納することによって、データ更新処理を

実現する。

【０１７９】

更に、記憶装置１－１では、上述の処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【０１８０】

また、自身が冗長データ更新処理を実施すると判断した記憶装置１－Ｎでは、前記ディスク制御部１４－Ｎを介して、ディスク部１１－Ｎに格納されている更新対象の冗長データ（旧冗長データ：DataPO）に対するディスクリード処理を開始する。

【０１８１】

更に、前記記憶装置１－１から受信した更新情報(Data1N')と上記旧冗長データ(DataPO)とから、例えば、排他的論理和演算結果等による新冗長データ(DataPN)を前記更新情報生成部１６－Ｎで生成し、該新冗長データ(DataPN)を、前記ディスク制御部１４－Ｎを介してディスク部１１－Ｎに格納することによって、冗長データ更新処理を実現する。

【０１８２】

更に、記憶装置１－Ｎでは、上述の処理に係るステータス情報と前段の装置から受信したステータス情報とに基づく新たなステータス情報を生成し、上位装置２に対して送信処理を実施した後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０１８３】

尚、上記冗長データの更新処理に当っては、前記Data Valid bit領域の対応するフラグが“有効”に設定されているか否かを検出し、“有効”に設定されている場合についてのみ冗長データの更新処理を実施し、“無効”の設定が残っている場合は受信したデータフレーム情報を後段の装置に対して再送することによって、例えば、冗長データ更新を実施する記憶装置の後段にデータ更新処理対象の記憶装置が位置する場合でも、上記データ更新処理／冗長データ更新処理を実現することが可能となる。

【０１８４】

(2-4) ライト処理（複数台の記憶装置でデータ更新処理実現）

図３１は、上位装置２からの前記アクセス要求が、ライトデータ要求で、且つ、複数台の記憶装置のデータ更新処理と一台の記憶装置の冗長データ更新処理とを実施する場合のデータ転送処理例を示す図であり、同図においては、記憶装置１－１／２がデータ更新処理(Data1O/2O→Data1N/2N)を、記憶装置１－Ｎが冗長データ更新処理(DataPO→DataPN)を実施するものとしている。

【０１８５】

前述と同様に、各記憶装置１－１～Ｎは、前記受信コマンド格納部２６に格納したアクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を自身が処理するか否か、更に、該アクセス要求に伴う冗長データ更新処理を実施するか否かを判別し、自身が処理しないと判断したディスク装置では、前記アクセス要求を削除することによって、上記ライトデータ要求に掛かる処理を完了する。

【０１８６】

自身がデータ更新処理を実施すると判断した記憶装置１－１では、前記ディスク制御部１４－１を介して、ディスク部１１－１に格納されている更新対象のデータ（旧データ：Data1O）に対するディスクリード処理を開始する。

【０１８７】

更に、上位装置２から前記インタフェース制御部１３－１を介して受信したデータ（新データ：Data1N/2N）のうち、自身の新データ(Data1N)と上記旧データ(Data1O)とから、例えば、排他的論理和演算結果等による更新情報(Data1N')を前記更新情報生成部１６－１で生成する。

【０１８８】

更に、前記上位装置２から受信した新データ(Data1N/2N)の内、自身に対する新データ(Data1N)を上記更新情報(Data1N')に置換え、当該更新情報を格納するセグメントに対応するフラグに“有効”、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記Data Valid bit領域を付加した前記データフレーム情報の送信処理を実施する。

【０１８９】

更に、記憶装置１－１では、前述と同様にデータ更新処理、並びに、ステータス送信処理等を実施し、一連のライトデータ要求処理を完了する。

【０１９０】

また、自身がデータ更新処理を実施すると判断した記憶装置１－２においても、前記ディスク制御部１４－２を介して、ディスク部１１－２に格納されている更新対象のデータ（旧データ：Data20）に対するディスクリード処理を開始する。

【０１９１】

更に、前段の記憶装置１－１から前記インタフェース制御部１３－２を介して受信したデータ（更新情報：Data1N'、新データ：Data2N）の内、自身に対する新データ(Data2N)と上記旧データ(Data20)とから、例えば、排他的論理和演算結果等による更新情報(Data2N')を前記更新情報生成部１６－２で生成する。

【０１９２】

更に、前段の記憶装置１－１から受信したデータ（Data1N'、Data2N）の内、自身に対する新データ(Data2N)を上記更新情報(Data2N')に置換え、当該更新情報を格納するセグメントに対応するフラグを“有効”に設定した前記データフレーム情報の送信処理を実施する。

【０１９３】

更に、前述と同様にデータ更新処理、並びに、前段の装置から受信したステータス情報と自身のステータス情報とから、新たなステータス情報の生成送信処理を実施し、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【０１９４】

また、自身が冗長データ更新処理を実施すると判断した記憶装置１－Ｎにおいても、前記ディスク制御部１４－Ｎを介して、ディスク部１１－Ｎに格納されている更新対象の冗長データ（旧冗長データ：DataP0）に対するディスクリード処理を開始する。

【０１９５】

更に、前記記憶装置１－２から受信した更新情報(Data1N'/2N')と上記旧冗長データ(DataP0)とから、例えば、排他的論理和演算結果等による新冗長データ(DataPN)を前記更新情報生成部１６－Ｎで生成し、該新冗長データ(DataPN)を、前記ディスク制御部１４－Ｎを介してディスク部１１－Ｎに格納することによって、冗長データ更新処理を実現する。

【０１９６】

更に、記憶装置１－Ｎでは、上述の処理に係るステータス情報と上述のデータ更新処理を実施した記憶装置１－１／２から受信したステータス情報とに基づく新たなステータス情報を生成し、上位装置２に対して送信処理を実施した後、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【０１９７】

尚、上記冗長データの更新処理に当っては、前述と同様に、前記Data Valid bit領域の対応するフラグの設定値に基づき実施するものである。

【０１９８】

次に、前記記憶装置サブシステム４を構成する前記記憶装置１における、上述のリードデータ要求処理とライトデータ要求処理とについて説明する。

【０１９９】

図３２は、上位装置２からのリードデータ要求に対する、記憶装置１での処理フロー例を示す図である。

【０２００】

記憶装置１は、前記協調制御情報２９と、前記受信コマンド格納部２６に格納したアクセス要求情報とに基づき、該アクセス要求を自身が処理するか否か（データ送信要否）の判別処理を実施する(ステップ100)。

【０２０１】

自身が処理しないと判断した記憶装置１は、後述のアクセス要求削除(ステップ106)を実施し、前記リードデータ要求に掛かる処理を完了する。

【０２０２】

自身が処理すると判断した記憶装置１は、前記ディスク制御部１４を介してデ

ィスク部１１から前記バッファ部１２の前記送受信データ格納部２７に対するディスクリード処理を開始し(ステップ101)、更に、前記協調制御情報２９と、前記アクセス要求情報とに基づき、自身がリードデータの送信開始を許可されたか否か（リードデータ送信開始可否検出）の判別処理を実施する(ステップ102)。

【０２０３】

自身がリードデータの送信の開始を許可された記憶装置１は、前記ユーザデータセグメント０に自身のリードデータを、他のユーザデータセグメントにはダミーデータを格納し、更に、前記ユーザデータセグメント０に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報を生成し、後段の装置に対する送信処理を実施する(ステップ103)。

【０２０４】

ここで、上記自身がリードデータの送信の開始を許可された記憶装置１以外に、前記上位装置２からのアクセス要求を協調して処理する記憶装置が存在しない場合、当該記憶装置１は、上記フラグ情報(Data Valid bit領域)の設定／追加を実施することなく、リードデータのみから構成されるデータフレーム情報を生成し、上位装置２に対する送信処理を実施する。

【０２０５】

一方、自身がリードデータの送信の開始を禁止された記憶装置１は、前段の記憶装置からの前記リードデータフレームの受信を待ち(ステップ107)、該リードデータフレームの受信を契機として(ステップ108)、受信したリードデータフレームの更新処理（自身のリードデータを格納すべきユーザデータセグメントに自身のリードデータを格納し、更に、当該ユーザデータセグメントに対応するフラグに“有効”を設定）を行い、更新したリードデータフレームの送信処理を実施する(ステップ109)。

【０２０６】

ここで、自身がリードデータの送信の開始を禁止された記憶装置１で、且つ、前記ユーザデータセグメントの最終セグメントに自身のリードデータを格納する記憶装置１は、上記フラグ情報(Data Valid bit領域)を削除し、リードデータの

みから構成されるデータフレーム情報を生成し、上位装置２に対する送信処理を実施する。

【０２０７】

また、自身が処理すると判断した記憶装置１は、前記協調制御情報２９と、前記受信コマンド格納部２６に格納したアクセス要求情報とに基づき、該アクセス要求処理の完了報告送信開始を許可されたか否か（完了報告送信開始可否検出）の判別処理を実施する(ステップ104)。

【０２０８】

自身が完了報告の送信の開始を許可された記憶装置１は、上記リードデータ処理に係るステータス情報を生成し、後段の装置に対する送信処理を実施し(ステップ105)、更に、ステータス情報の送信処理後、前記アクセス要求を削除する(ステップ106)ことによって、一連のリードデータ要求処理を完了する。

【０２０９】

一方、自身が完了報告の送信の開始を禁止された記憶装置１は、前段の記憶装置からの前記ステータス情報の受信を待ち(ステップ110)、該ステータス情報の受信を契機として(ステップ111)、受信したステータス情報の更新処理（自身のリードデータ処理結果の反映）を行い、更新したステータス情報の送信処理を実施(ステップ112)後、前記アクセス要求を削除する(ステップ106)ことによって、一連のリードデータ要求処理を完了する。

【０２１０】

尚、上述の実施例において、前記ユーザデータセグメント０に自身のリードデータを格納するか否かを前記リードデータ送信開始可否検出の条件例とし、更に、前記ユーザデータセグメント０に自身のリードデータを格納する記憶装置において、前記フラグ情報(Data Valid bit領域)を付加した前記データフレーム情報を生成するものとして説明したが、これに限定するものではない。同様に、前記ユーザデータセグメントの最終セグメントに自身のリードデータを格納する記憶装置１において、上記フラグ情報(Data Valid bit領域)を削除し、リードデータから構成されるデータフレーム情報を生成するものとして説明したが、これに限定するものではなく、任意のアルゴリズムで上記記憶装置の決定を行うことが可

能である。

【０２１１】

図３３は、上位装置２からのライトデータ要求に対する、記憶装置１での処理フロー例を示す図である。

【０２１２】

記憶装置１は、前記協調制御情報２９と、前記受信コマンド格納部２６に格納したアクセス要求情報とに基づき、該アクセス要求を自身が処理（データ更新）するか否か、更に、該アクセス要求に伴う冗長データ更新処理を実施するか否かの判別処理を実施する(ステップ200)。

【０２１３】

自身が処理しないと判断した記憶装置１は、後述のアクセス要求削除(ステップ209)を実施し、前記ライトデータ要求に掛かる処理を完了する。

【０２１４】

自身がデータ更新処理を実施すると判断した記憶装置１は、前記ディスク制御部１４を介して、ディスク部１１に格納されている更新対象のデータ（旧データ）に対するディスクリード処理を開始する(ステップ201)。

【０２１５】

更に、前段の装置からライトデータフレーム情報の受信を待ち(ステップ202)、該ライトデータフレーム情報の受信を契機として(ステップ203)、該受信ライトデータフレーム情報を構成するライトデータ（新データ集合体）の内、自身が格納すべき新データと上記旧データとから、排他的論理和演算等による更新情報を前記更新情報生成部１６で生成する(ステップ204)。

【０２１６】

更に、上記受信したライトデータフレーム情報を構成するライトデータ（新データ集合体）の内、自身に対する新データを上記更新情報に置換え、当該更新情報を格納するユーザデータセグメントに対応するフラグを“有効”に設定した前記ライトデータフレーム情報の送信処理を実施する(ステップ205)。

【０２１７】

ここで、上記受信したライトデータフレーム情報が、前記フラグ情報(Data Va

lid bit領域)を付加していない場合、記憶装置１は、自身の更新情報を格納するユーザデータセグメントに対応するフラグに“有効”を、 他のユーザデータセグメントに対応するフラグに“無効”を設定した上記フラグ情報を付加したライトデータフレーム情報を生成し、当該ライトデータフレーム情報の送信処理を実施する。

【０２１８】

更に、上記自身が格納すべき新データを、前記ディスク制御部１４を介してディスク部１１に格納（ディスクライト）することによって、ライトデータ更新処理を実施する(ステップ206)。

【０２１９】

また、自身が冗長データ更新処理を実施すると判断した記憶装置１は、前記ディスク制御部１４を介して、ディスク部１１に格納されている更新対象の冗長データ（旧冗長データ）に対するディスクリード処理を開始する(ステップ210)。

【０２２０】

更に、前段の装置からライトデータフレーム情報の受信を待ち(ステップ211)、該ライトデータフレーム情報を構成する前記ユーザデータセグメントに対応する全ての前記フラグ情報(Data Valid bit領域)が“有効”に設定されていることの検出を契機として(ステップ212)、該ライトデータフレーム情報を構成する更新情報（集合体）と上記旧冗長データとから、排他的論理和演算等による新冗長データを前記更新情報生成部１６で生成する(ステップ213)。なお、ここでは、条件として、全てのフラグ情報としたが、所定の条件を満たすフラグ情報としてもよい。該新冗長データを、前記ディスク制御部１４を介してディスク部１１に格納することによって、冗長データ更新処理を実現する(ステップ214)。

【０２２１】

ここで、上記ライトデータフレーム情報を構成する前記ユーザデータセグメントに対応する全ての前記フラグ情報が“有効”に設定されていない場合、受信したライトデータフレームを再送するものとする。

【０２２２】

これによって、データ更新対象の記憶装置が、上記冗長データ更新対象記憶装

置の後段に位置する場合でも、上記データ更新対象の記憶装置による前記更新情報の生成／送信処理が実現可能となる。

【０２２３】

尚、当該処理によって、再度、上記ライトデータフレーム情報を受信した記憶装置は、該ライトデータフレーム情報を構成する前記フラグ情報に基づき、例えば、自身が格納すべきユーザデータセグメントに対応するフラグが“有効”に設定されていることを条件として、再送処理を実施する。

【０２２４】

また、自身がデータ更新処理、或いは、冗長データ更新処理を実施すると判断した記憶装置１は、前記協調制御情報２９と、前記受信コマンド格納部２６に格納したアクセス要求情報とに基づき、該アクセス要求処理の完了報告送信開始を許可されたか否か（完了報告送信開始可否検出）の判別処理を実施する(ステップ207)。

【０２２５】

自身が完了報告の送信の開始を許可された記憶装置１は、上記ライトデータ処理に係るステータス情報を生成し、後段の装置に対する送信処理を実施し(ステップ208)、更に、ステータス情報の送信処理後、前記アクセス要求を削除する(ステップ209)ことによって、一連のライトデータ要求処理を完了する。

【０２２６】

一方、自身が完了報告の送信の開始を禁止された記憶装置１は、前段の記憶装置からの前記ステータス情報の受信を待ち(ステップ215)、該ステータス情報の受信を契機として(ステップ216)、受信したステータス情報の更新処理（自身のライトデータ処理結果の反映）を行い、更新したステータス情報の送信処理を実施(ステップ217)後、前記アクセス要求を削除する(ステップ209)ことによって、一連のライトデータ要求処理を完了する。

【０２２７】

尚、上述の実施例において、データ更新処理対象の全ての記憶装置でのデータ更新処理実施後、冗長データの更新処理を行うこととしていることから、上記冗長データ更新対象の記憶装置が最終的なステータス情報を生成／送信するものと

しているが、これに限定するものではない。同様に最終的なステータス情報の生成方法についても、個々の記憶装置がステータス情報を更新する方法を例に説明したが、これに限定するものではなく、例えば、前記記憶装置サブシステム４を構成する特定の記憶装置が、上記ステータス情報の生成送信を管理することも可能である。

【０２２８】

［第五の実施形態］　（縮退動作）

以下、本発明に係る第五の実施形態を、図３４～１７を用いて説明する。

【０２２９】

第五の実施形態は、第四の実施形態の記憶装置が、更に、協調して処理する他の記憶装置の動作状態情報を追加保持することによって、前記記憶装置サブシステムを構成する記憶装置に障害が発生した場合でも、上位装置から記憶装置サブシステムに対するアクセス要求処理を実現するものである。

【０２３０】

図３４は、前述の実施形態の説明で用いた前記協調制御情報２９の冗長構成管理情報を構成する情報として、ＲＡＩＤレベル情報に“５”、データディスク台数／冗長データディスク台数情報に“３／１”が設定され、更に、動作モード情報と障害記憶装置識別情報（障害記憶装置ＩＤ）とを保持し、ID_2の記憶装置障害による縮退モードが設定されている例である。

【０２３１】

尚、本実施形態における記憶装置は、前述の構成と同じであることから図を省略する。

【０２３２】

また、後述のデータ復元処理が不要な場合は、前述の実施形態と同一の制御で実施可能であることから説明を省略する。

【０２３３】

以下、記憶装置サブシステム４を構成する記憶装置１－２おいて障害が発生した場合を例に、前記上位装置２から記憶装置サブシステム４に対するアクセス要求を、上記記憶装置１－２を除く他の記憶装置が協調して処理する動作を説明す

る。

【０２３４】

(1) 障害記憶装置に対するリード処理（複数台の記憶装置でデータ復元処理実現）

図３５は、障害状態にある記憶装置１－２に格納されたデータに対する上位装置２からのリードデータ要求を、上記記憶装置１－２以外の記憶装置で処理する場合のデータ転送処理例を示す図である。

【０２３５】

前記受信コマンド格納部２６に上記アクセス要求情報を格納した記憶装置１－２以外の記憶装置は、上記アクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を処理するに当り、障害状態にある記憶装置１－２に格納されたデータの復元処理が必要であることを検出し、更に、自身が該データ復元処理を実施するか否かを判別する。

【０２３６】

障害状態にある記憶装置に格納されたデータの復元処理が必要と判断した記憶装置は、必要ならば、前記ディスク制御部１４を介して、ディスク部１１に格納されている上記データ復元処理に必要なデータ（図中、Data1, Data3, Data4）に対するディスクリード処理を開始する。

【０２３７】

尚、以下の説明では、前記アクセス要求情報と前記協調制御情報２９とに基づき、最前段に位置する（上位装置からの情報を最も早く受信する）記憶装置が、上記データ復元処理に係るデータの送信を開始し、最後段に位置する（上位装置からの情報を最後に受信する）記憶装置が、上記データ復元処理を実施する場合を例に説明するが、これに限るものではない。

【０２３８】

自身が前記データ復元処理に係るデータ送信の開始を許可された記憶装置１－１では、例えば、前記ユーザデータセグメント０に自身のデータを格納し、更に、前記ユーザデータセグメント０に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報

の送信処理を実施し、更に、該処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０２３９】

また、自身が前記データ復元処理に係るデータ送信の開始を禁止された記憶装置１－３では、上記記憶装置１－１から受信した上記データフレーム情報に、自身のデータを格納したユーザデータセグメント１を追加し、更に、前記ユーザデータセグメント１に対応するフラグに“有効”を設定したデータフレーム情報の生成／送信処理を実施し、更に、該処理に係るステータス情報と前段の装置から受信したステータス情報とに基づく新たなステータス情報を生成送信後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０２４０】

更に、自身が前記データ復元処理を実施すると判断した記憶装置１－４では、上記記憶装置１－１／３から受信したデータ（Data1，Data3）と、自身が格納し上記データ復元処理に必要なデータ(Data4)とから、前記更新情報生成部１６－４での排他的論理和演算等によって、前記障害状態にある記憶装置１－２に格納されたデータ(Data2)の復元処理を実施し、更に、復元したデータ(Data2)を用いて前記上位装置２に対してリードデータフレーム情報の送信処理を実施する。

【０２４１】

尚、上記データ復元処理を実施する最終段の記憶装置、即ち、上位装置に対する最終形のデータフレーム情報を生成する記憶装置においては、前記ユーザデータセグメントのデータが有効か否かを示すフラグ(Data Valid bit)領域を削除して、ユーザデータのみから構成されるペイロードとして上位装置に送信することが望ましい。

【０２４２】

また、上位装置２に対するステータス情報の送信処理としては、上記データ送信処理を実施する最終段の記憶装置、即ち、上位装置に対する最終形のデータフレーム情報を生成する記憶装置において、前段の装置から受信したステータス情報と自身のステータス情報とから新たなステータス情報を生成送信することによって、上位装置２に対するステータス情報の生成を実現している。

【０２４３】

また、上述の実施形態では、障害状態にある記憶装置１－２に格納されたデータのみに対する上位装置２からのリードデータ要求を、上記記憶装置１－２以外の記憶装置で処理する場合を例に説明したが、これに限定するものではなく、例えば、記憶装置１－１／２に格納されたデータに対するリードデータ要求の場合、上記記憶装置１－４において、前記記憶装置１－１から受信したデータ(Data1)と自身が復元したデータ(Data2)とを用いて、上位装置に対する最終形のデータフレーム情報を生成／送信することが可能である。

【０２４４】

(2) 障害記憶装置に対するライト処理（複数台の記憶装置で冗長データ更新処理実現）

図３６は、障害状態にある記憶装置１－２に対する上位装置２からのライトデータ要求を、上記記憶装置１－２以外の記憶装置で処理する場合のデータ転送処理例を示す図である。

【０２４５】

前記受信コマンド格納部２６に上記アクセス要求情報を格納した記憶装置１－２以外の記憶装置は、上記アクセス要求情報と前記協調制御情報２９とに基づき、データ更新処理を自身が実施するか否か、或いは、該アクセス要求に伴う冗長データ更新処理を自身が実施するか否かを判別する。

【０２４６】

上記データ更新処理、或いは、冗長データ更新処理を実施しないと判断した記憶装置は、必要ならば、前記ディスク制御部１４を介して、ディスク部１１に格納されている上記冗長データ更新処理に必要なデータ（図中、Data1, Data3）に対するディスクリード処理を開始する。

【０２４７】

尚、以下の説明では、最前段に位置する（上位装置からの情報を最も早く受信する）記憶装置が、上記冗長データ更新処理に係るデータの送信を開始し、最後段に位置する（上位装置からの情報を最後に受信する）記憶装置が、上記データ復元処理を実施する場合を例に説明するが、これに限るものではない。

【０２４８】

自身が前記冗長データ更新処理に係るデータ送信の開始を許可された記憶装置１－１では、例えば、自身のデータ(Data1)を前記ユーザデータセグメント０に、上位装置２から受信した新データ(Data2)を前記ユーザデータセグメント１格納し、更に、前記ユーザデータセグメント０／１に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報の送信処理を実施し、更に、該処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【０２４９】

また、自身が前記冗長データ更新処理に係るデータ送信の開始を禁止された記憶装置１－３では、上記記憶装置１－１から受信した上記データフレーム情報に、自身のデータ(Data3)を格納したユーザデータセグメント２を追加し、更に、前記ユーザデータセグメント２に対応するフラグに“有効”を設定したデータフレーム情報の生成／送信処理を実施し、更に、該処理に係るステータス情報と前段の装置から受信したステータス情報とに基づく新たなステータス情報を生成送信後、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了する。

【０２５０】

更に、自身が前記冗長データ更新処理を実施すると判断した記憶装置１－４では、最終的に上記記憶装置１－３から受信したデータ（Data1，Data2，Data3）から、前記更新情報生成部１６－４での排他的論理和演算等によって、新たな冗長データ(DataP)を生成し、更に、生成した冗長データ(DataP)を、前記ディスク制御部１４－４を介してディスク部１１－４に格納することによって、冗長データ更新処理を実現する。

【０２５１】

更に、記憶装置１－４では、上述の処理に係るステータス情報と前段の装置から受信したステータス情報とに基づく新たなステータス情報を生成し、上位装置２に対して送信処理を実施した後、前記アクセス要求を削除することによって、

一連のライトデータ要求処理を完了する。

【０２５２】

尚、上記冗長データの更新処理に当っては、前記Data Valid bit領域の対応するフラグが“有効”に設定されているか否かを検出し、“有効”に設定されている場合についてのみ冗長データの更新処理を実施し、“無効”の設定が残っている場合は受信したデータフレーム情報を後段の装置に対して再送することによって、例えば、冗長データ更新を実施する記憶装置の後段にデータ更新処理対象の記憶装置が位置する場合でも、上記冗長データ更新処理を実現することが可能となる。

【０２５３】

また、上述の実施形態では、障害状態にある記憶装置１－２に格納されたデータのみに対する上位装置２からのライトデータ要求を、上記記憶装置１－２以外の記憶装置で処理する場合を例に説明したが、これに限定するものではなく、例えば、記憶装置１－１／２に格納されたデータに対するライトデータ要求の場合、上記記憶装置１－１において、上述のディスクリード処理を実施することなく、上位装置２から受信した新データ(Data1, Data2)の内、自身に対する新データ(Data1)を、前記ディスク制御部１４－１を介してディスク部１１－１に格納することによって、データ更新処理を実現し、更に、上位装置２から受信した新データ(Data1, Data2)を前記ユーザデータセグメント０／１格納し、更に、前記ユーザデータセグメント０／１に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報の送信処理を実施し、更に、該処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のライトデータ要求処理を完了することが可能である。

【０２５４】

更に、上記協調制御情報２９においては、上位装置から設定可能であることが望ましく、更に、自身の判断によって、前記動作モード情報等を設定可能であることが望ましい。

【０２５５】

［第六の実施形態］　（マルチパス制御）

以下、本発明に係る第六の実施形態を、図３７、１９を用いて説明する。

【０２５６】

第六の実施形態は、上述の実施形態の記憶装置が、更に、複数のインタフェース制御部を具備し、上位装置からのアクセス要求情報と、記憶装置サブシステムを構成する記憶装置間で転送されるユーザデータとの転送処理を、異なるインタフェース（伝送路）を用いて実現するものである。

【０２５７】

図３７は、複数(N，N>1)台の記憶装置１から構成される記憶装置サブシステム４と、上位装置２とから構成される情報処理システムの一例を示す図である。

【０２５８】

同図において、記憶装置サブシステム４を構成する個々の記憶装置１が、複数のインタフェース制御部Ａ／Ｂ（１３－ａ／ｂ）を具備し、該インタフェース制御部Ａ１３－ａを用いて、前述の実施形態と同様に、前記上位装置２と巡回型のインタフェース３－ａを介して直接接続される。同時に、該インタフェース制御部Ｂ１３－ｂを用いて、記憶装置サブシステム４を構成する個々の記憶装置１が、巡回型のインタフェース３－ｂを介して接続されている。

【０２５９】

尚、上記情報処理システムの構成例として、記憶装置サブシステム４を構成する個々の記憶装置１のみを、前記インタフェース制御部Ｂ１３－ｂを用いて、巡回型のインタフェース３－ｂを介して接続する構成としているが、後述の通り、これに限定するものではない。

【０２６０】

以下、記憶装置サブシステム４を構成する記憶装置１－２おいて障害が発生した状態で、上位装置２から上記障害状態にある記憶装置１－２に格納されたデータに対するリードデータ要求を、上記記憶装置１－２を除く他の記憶装置が協調して処理する場合を例に動作を説明する。

【０２６１】

図３８は、障害状態にある記憶装置１－２に格納されたデータに対する上位装

置２からのリードデータ要求を、上記記憶装置１－２以外の記憶装置で処理する場合のデータ転送処理例を示す図である。

【０２６２】

前記受信コマンド格納部２６に、インタフェース制御部Ａ１３－ａを介して、上記アクセス要求情報を格納した記憶装置１－２以外の記憶装置は、前述の実施形態と同様に、上記アクセス要求情報と前記協調制御情報２９とに基づき、該アクセス要求を処理するに当り、障害状態にある記憶装置１－２に格納されたデータの復元処理が必要であることを検出し、更に、自身が該データ復元処理を実施するか否かを判別する。

【０２６３】

障害状態にある記憶装置に格納されたデータの復元処理が必要と判断した記憶装置は、必要ならば、前記ディスク制御部１４を介して、ディスク部１１に格納されている上記データ復元処理に必要なデータ（図中、Data1，Data3，Data4）に対するディスクリード処理を開始する。

【０２６４】

自身が前記データ復元処理に係るデータ送信の開始を許可された記憶装置１－１では、例えば、前記ユーザデータセグメント０に自身のデータ(Data1)を格納し、更に、前記ユーザデータセグメント０に対応するフラグに“有効”を、他のユーザデータセグメントに対応するフラグに“無効”を設定した前記データフレーム情報の送信処理を、上記アクセス要求情報を受信したインタフェース制御部Ａとは異なるインタフェース制御部Ｂを用いて実施し、更に、該処理に係るステータス情報の生成送信後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０２６５】

また、自身が前記データ復元処理に係るデータ送信の開始を禁止された記憶装置１－３では、上記記憶装置１－１から受信した上記データフレーム情報に、自身のデータを格納したユーザデータセグメント１を追加し、更に、前記ユーザデータセグメント１に対応するフラグに“有効”を設定したデータフレーム情報の生成／送信処理、更に、該処理に係るステータス情報と前段の装置から受信した

ステータス情報とに基づく新たなステータス情報を生成送信処理を、上記記憶装置１－１と同様に、上記アクセス要求情報を受信したインタフェース制御部Ａとは異なるインタフェース制御部Ｂを用いて実施した後、前記アクセス要求を削除することによって、一連のリードデータ要求処理を完了する。

【０２６６】

更に、自身が前記データ復元処理を実施すると判断した記憶装置１－４では、上記記憶装置１－１／３から、前記インタフェース制御部Ｂ１３－４ｂを介して受信したデータ（Data1，Data3）と、自身が格納し上記データ復元処理に必要なデータ(Data4)とから、前記更新情報生成部１６－４での排他的論理和演算等によって、前記障害状態にある記憶装置１－２に格納されたデータ(Data2)の復元処理を実施し、更に、復元したデータ(Data2)を、前記インタフェース制御部Ａ１３－４ａ用いて前記上位装置２に対するリードデータフレーム情報の送信処理を実施するものである。

【０２６７】

尚、詳細な説明は省略するが、ライトデータ処理においても同様に、上位装置からのアクセス要求情報を受信したインタフェースと異なるインタフェースを用いて、記憶装置サブシステムを構成する記憶装置間で協調したデータ転送処理を実現することが可能である。

【０２６８】

また、上述の説明において、上位装置からのアクセス要求情報を受信したインタフェースと異なるインタフェースを用いて、記憶装置サブシステムを構成する記憶装置間で協調したデータ転送処理を実施する場合を例に説明しているが、これに限定するものではなく。例えば、フレーム送信時のインターフェースの稼動状況に基づき、前記協調処理を実施する記憶装置間でのデータ転送処理を行うインタフェースを選択することも可能であり、該データ転送処理を実施するインタフェース上に上位装置２が接続されていても何ら問題は無い。

【０２６９】

以上説明した各実施形態には、以下の効果がある。

本発明の第一の実施形態によれば、複数の記憶装置から構成される記憶装置シ

ステムにおいて、該記憶装置システムを構成する個々の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を、前記複数の記憶装置で共有する手段と、上記共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべきと判断した記憶装置に転送する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要で、また、記憶装置の利用効率の向上が期待される記憶装置システムが実現可能となる。

【０２７０】

本発明の第二の実施形態によれば、複数の記憶装置から構成される記憶装置システムにおいて、該記憶装置システムを構成する少なくとも一台の記憶装置が、上位装置から記憶装置システムに対するアクセス要求を受領し、更に、当該記憶装置において、上記アクセス要求を処理すべき記憶装置の判別を実施し、必要な場合、前記上位装置から記憶装置システムに対するアクセス要求を、処理すべき記憶装置に対して転送する手段を具備することによって、個々の記憶装置を制御する制御装置が不要で、また、記憶装置の利用効率の向上が期待される記憶装置システムが実現可能となる。

【０２７１】

本発明の第三の実施形態によれば、複数の記憶装置から構成される記憶装置システムにおいて、記憶装置システムを構成する個々の記憶装置が、少なくとも上位装置から記憶装置システムに対するライトデータ要求を共有する手段と、上記共有したライトデータ要求について、対を成す複数の記憶装置間でライトデータの転送処理と、当該ライトデータ要求に対するステータス情報の送信処理とを同期して実施する手段とを具備することによって、個々の記憶装置を制御する制御装置が不要でインタフェース負荷を軽減する記憶装置システムが実現可能となる。

【０２７２】

【発明の効果】

本発明により、複数の記憶装置から構成されるRAID1を含む記憶装置システムにおいて、上位装置からのアクセス要求を、記憶装置システムを構成する複数の

記憶装置間で協調して処理ことが可能となる。

【図面の簡単な説明】

【図１】本発明の情報処理システムの一構成例を示す図

【図２】本発明の第一実施形態に係る記憶装置の一構成例を示す図

【図３】本発明の第一実施形態に係るリード処理フローの一例を示す図

【図４】本発明の第一実施形態に係るリード処理における情報転送の一例を示す図

【図５】本発明の第一実施形態に係るライト処理フローの一例を示す図

【図６】本発明の第一実施形態に係るライト処理における情報転送の一例を示す図

【図７】本発明の第一実施形態に係る未更新領域管理情報の一構成例を示す図

【図８】本発明の第一実施形態に係る未更新領域の更新処理における情報転送の一例を示す図

【図９】本発明の第二実施形態に係る記憶装置の一構成例を示す図

【図１０】本発明の第二実施形態に係るリード処理フローの一例を示す図

【図１１】本発明の第二実施形態に係るリード処理における情報転送の一例を示す図

【図１２】本発明の第二実施形態に係るライト処理フローの一例を示す図

【図１３】本発明の第二実施形態に係る未更新領域の更新処理における情報転送の一例を示す図

【図１４】本発明の第三実施形態に係るライト処理フローの一例を示す図

【図１５】本発明の第三実施形態に係るライト処理における情報転送の一例を示す図

【図１６】本発明の協調制御情報の一構成例を示す図

【図１７】アクセス要求情報の一構成例を示す図

【図１８】フレームヘッダ情報の一構成例を示す図

【図１９】本発明のライト処理管理情報の一構成例を示す図

【図２０】本発明の情報処理システムの一構成例を示す図

【図２１】本発明の第四実施形態に係る記憶装置の一構成例を示す図

【図２２】本発明の第四実施形態に係る協調制御情報の構成及び設定例を示す図

【図２３】本発明で使用されるフレーム情報の一構成例を示す図

【図２４】本発明で使用されるフレーム情報を構成するフレーム情報の一構成例を示す図

【図２５】本発明で使用されるフレーム情報を構成するアクセス要求情報の一構成例を示す図

【図２６】本発明におけるアクセス要求情報の転送処理フローの一例を示す図

【図２７】本発明の第四実施形態に係るリードデータ転送処理（単一）の一例を示す図

【図２８】本発明の第四実施形態に係るリードデータ転送処理（複数）の一例を示す図

【図２９】本発明におけるデータフレーム情報の一構成例を示す図

【図３０】本発明の第四実施形態に係るライトデータ転送処理（単一）の一例を示す図

【図３１】本発明の第四実施形態に係るライトデータ転送処理（複数）の一例を示す図

【図３２】本発明の第四実施形態に係るリード処理フローの一例を示す図

【図３３】本発明の第四実施形態に係るライト処理フローの一例を示す図

【図３４】本発明の第五実施形態に係る協調制御情報の構成及び設定例を示す図

【図３５】本発明の第五実施形態に係るリードデータ転送処理（縮退）の一例を示す図

【図３６】本発明の第五実施形態に係るライトデータ転送処理（縮退）の一例を示す図

【図３７】本発明の第六実施形態に係る情報処理システムの一構成例を示す図

【図３８】本発明の第六実施形態に係るリードデータ転送処理（縮退）の一例を示す図

【符号の説明】

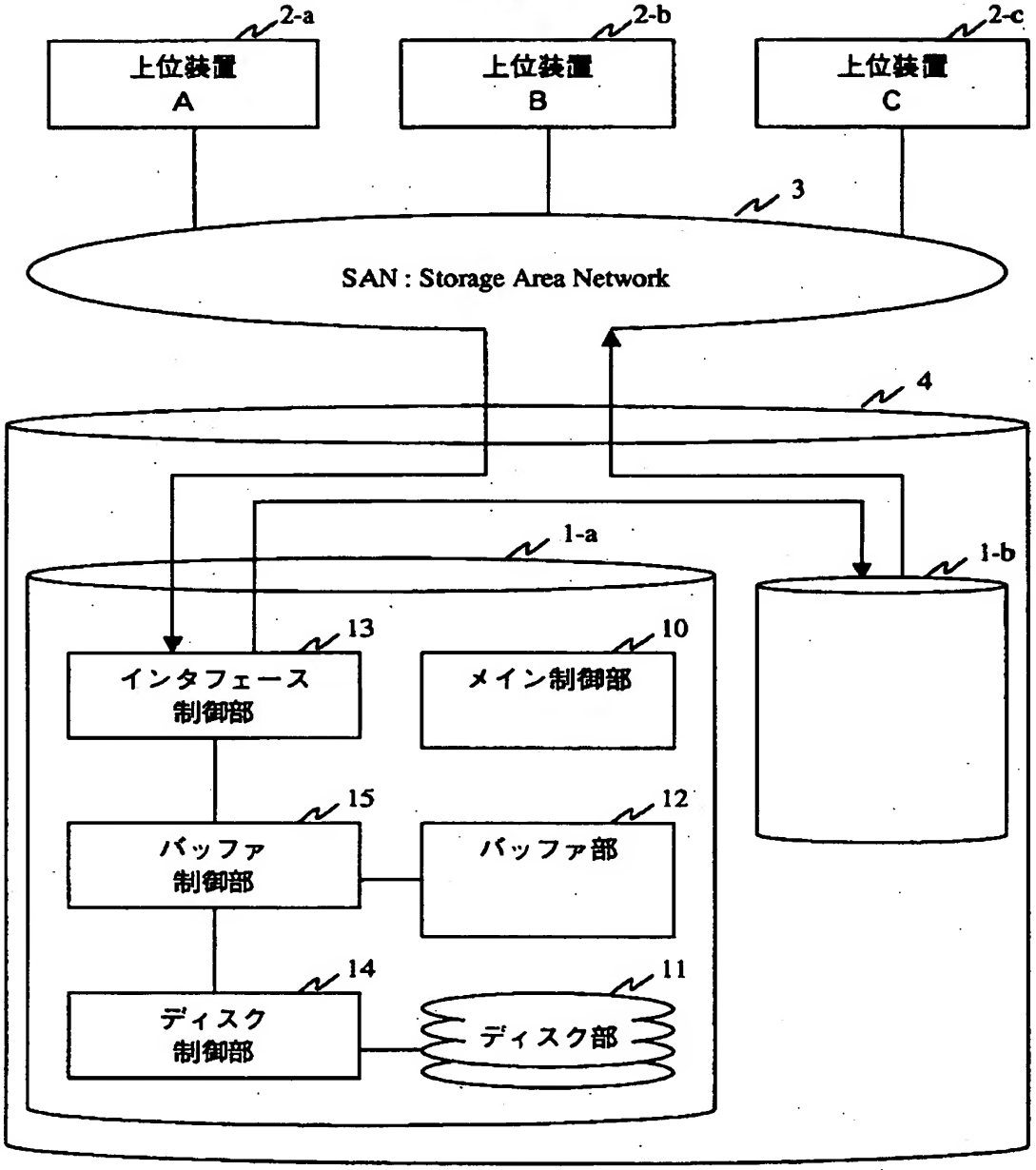
１・・・記憶装置、２・・・上位装置、３・・・インタフェース（SAN：Storage Area Network）、４・・・記憶装置システム、１０・・・メイン制御部、１１

・・・ディスク部（記憶媒体）、１２・・・バッファ部、１３・・・インタフェース制御部、１４・・・ディスク制御部、１５・・・バッファ制御部、２０・・・受信部、２１・・・送信部、２２・・・受信フレーム処理部、２３・・・送信フレーム生成部、２４・・・再送判定部、２５・・・出力選択部、２６・・・受信コマンド格納部、２７・・・送受信データ格納部、２８・・・協調制御情報、２９・・・コマンド転送制御部、３０・・・未更新領域管理情報

【書類名】　図面

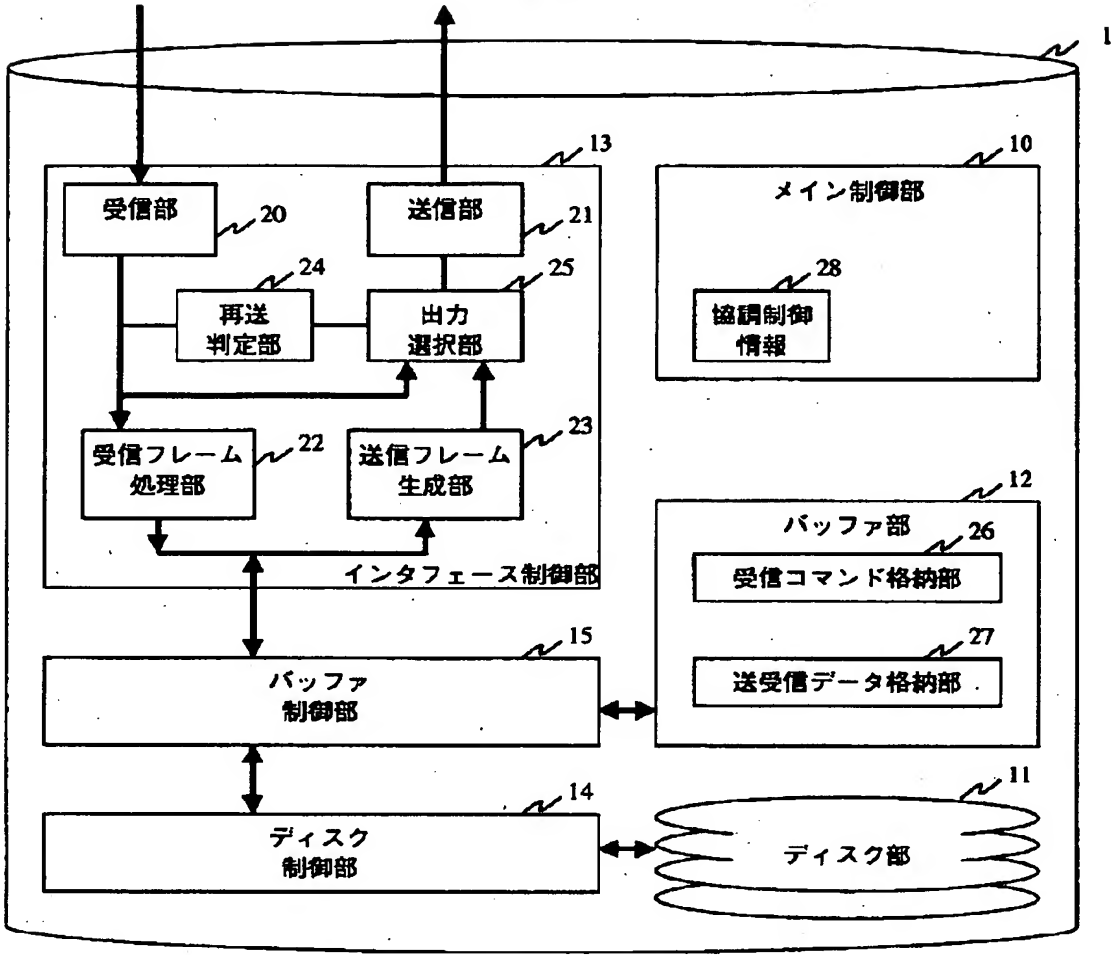
【図１】

図１


2-a
上位装置
A
2-b
上位装置
B
2-c
上位装置
C
3
SAN : Storage Area Network
4
1-a
1-b
13
インタフェース
制御部
10
メイン制御部
15
バッファ
制御部
12
バッファ部
14
ディスク
制御部
11
ディスク部

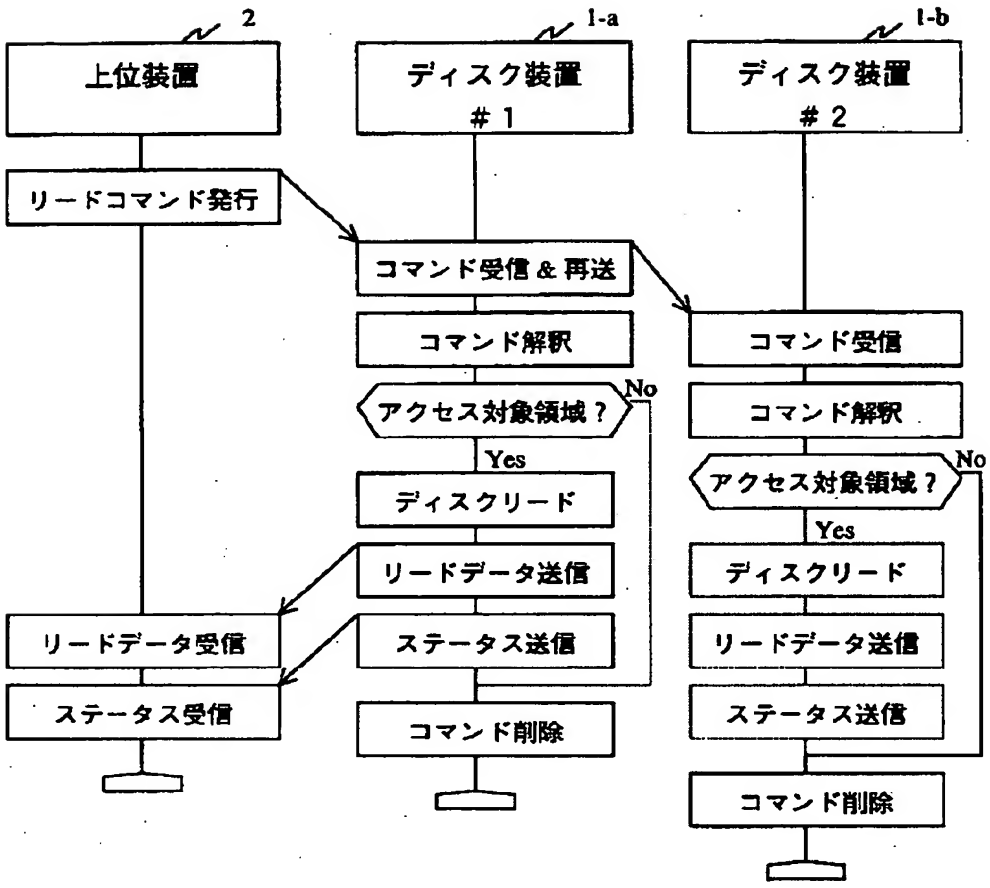
【図２】

図２


1
13
受信部
20
送信部
21
24
再送
判定部
25
出力
選択部
受信フレーム
処理部
22
送信フレーム
生成部
23
インタフェース制御部
10
メイン制御部
28
協調制御
情報
12
バッファ部
26
受信コマンド格納部
27
送受信データ格納部
15
バッファ
制御部
14
ディスク
制御部
11
ディスク部

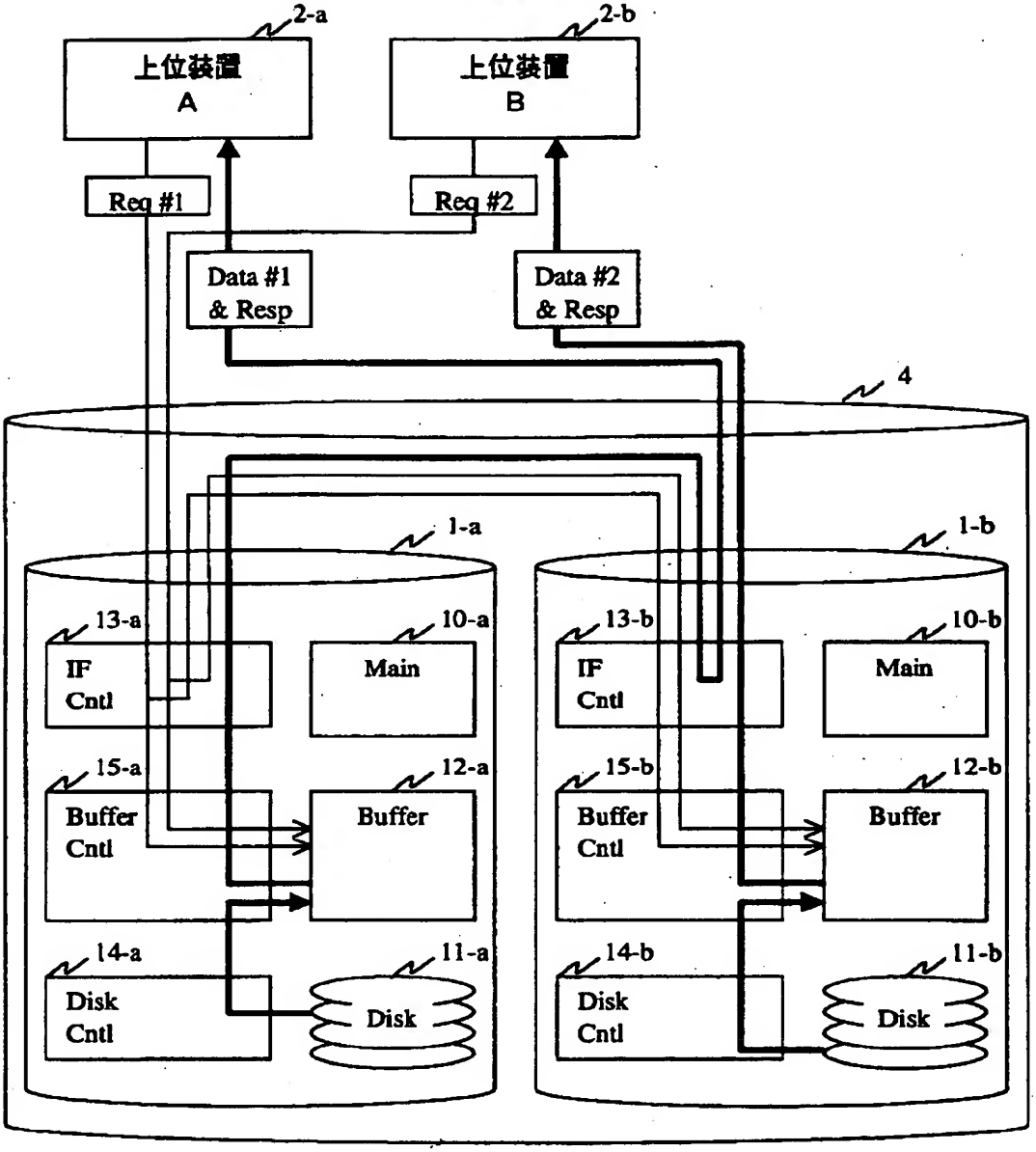
【図３】

図３


2
1-a
1-b
上位装置
ディスク装置
＃１
ディスク装置
＃２
リードコマンド発行
コマンド受信＆再送
コマンド解釈
コマンド受信
アクセス対象領域？
No
コマンド解釈
Yes
ディスクリード
アクセス対象領域？
No
Yes
リードデータ送信
ディスクリード
リードデータ受信
ステータス送信
リードデータ送信
ステータス受信
コマンド削除
ステータス送信
コマンド削除

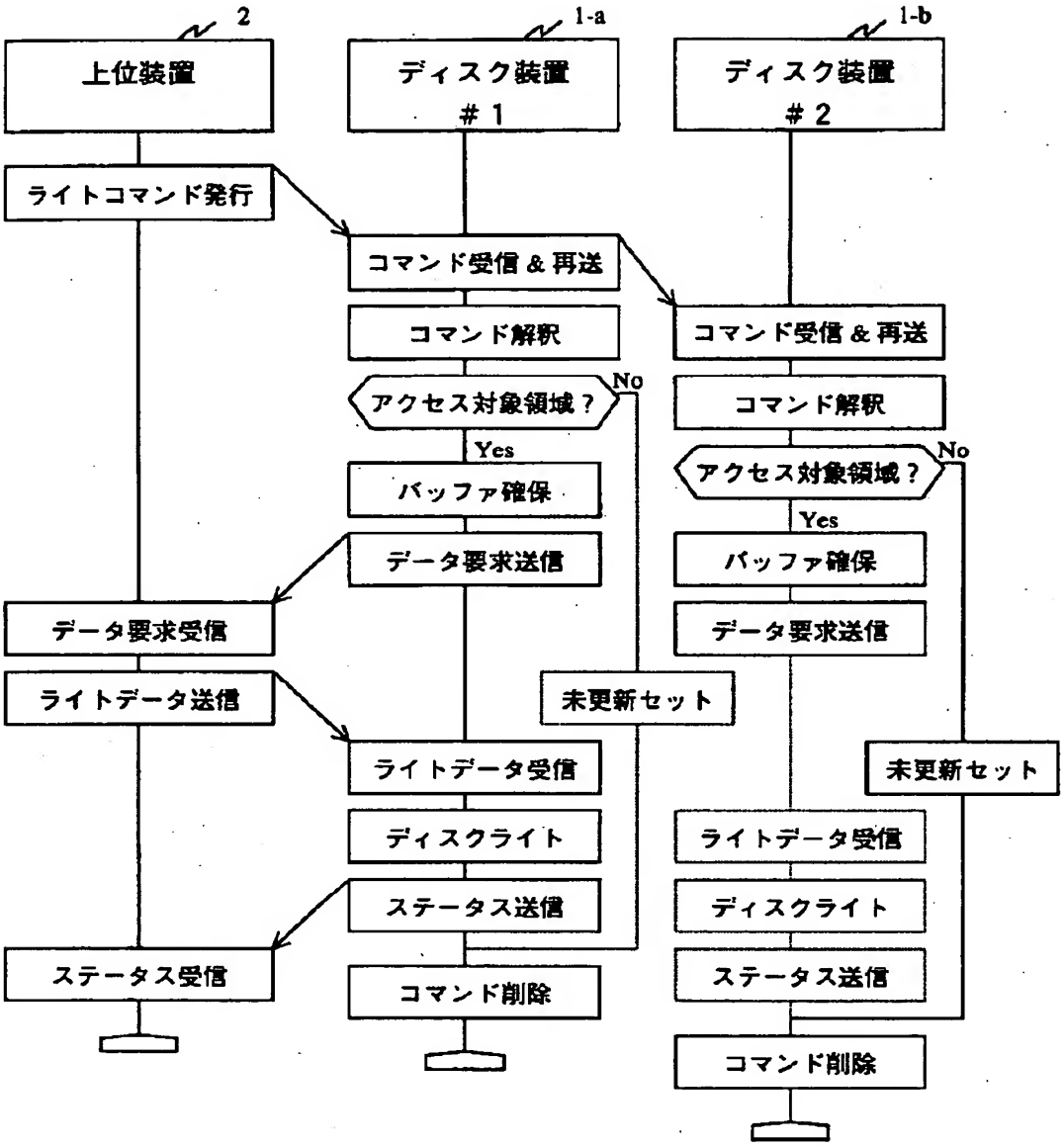
【図４】

図４


2-a
上位装置
A
2-b
上位装置
B
Req #1
Req #2
Data #1
& Resp
Data #2
& Resp
4
1-a
1-b
13-a
IF
Cntl
10-a
Main
13-b
IF
Cntl
10-b
Main
15-a
Buffer
Cntl
12-a
Buffer
15-b
Buffer
Cntl
12-b
Buffer
14-a
Disk
Cntl
11-a
Disk
14-b
Disk
Cntl
11-b
Disk

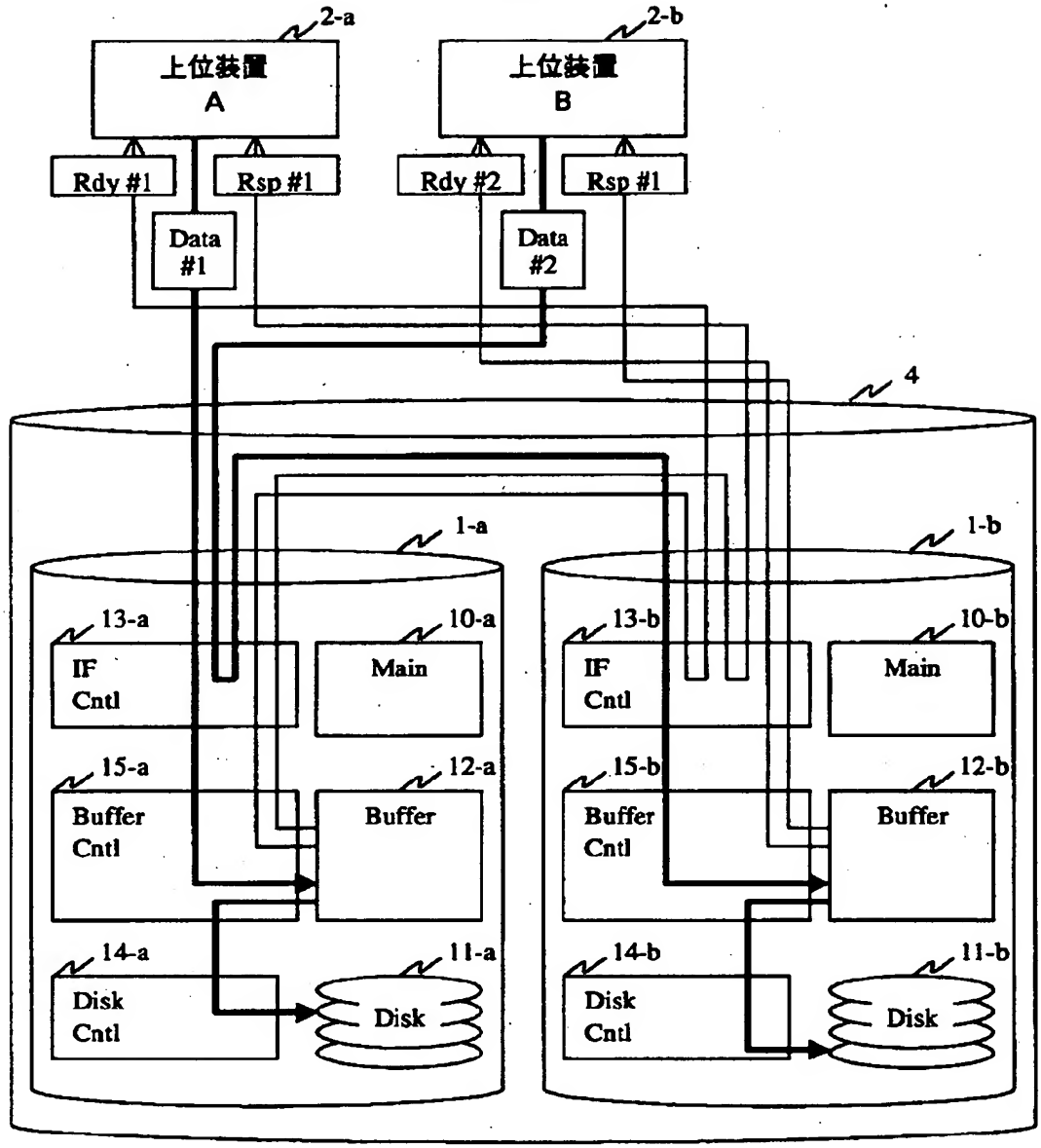
【図５】

図５


2
1-a
1-b
上位装置
ディスク装置
＃１
ディスク装置
＃２
ライトコマンド発行
コマンド受信＆再送
コマンド解釈
コマンド受信＆再送
アクセス対象領域？
No
コマンド解釈
Yes
バッファ確保
アクセス対象領域？
No
Yes
データ要求送信
バッファ確保
データ要求受信
データ要求送信
ライトデータ送信
未更新セット
ライトデータ受信
未更新セット
ディスクライト
ライトデータ受信
ステータス送信
ディスクライト
ステータス受信
ステータス送信
コマンド削除
コマンド削除

【図６】

図６


2-a
2-b
上位装置
A
上位装置
B
Rdy #1
Rsp #1
Rdy #2
Rsp #1
Data
#1
Data
#2
4
1-a
1-b
13-a
10-a
IF
Cntl
Main
13-b
10-b
IF
Cntl
Main
15-a
12-a
Buffer
Cntl
Buffer
15-b
12-b
Buffer
Cntl
Buffer
14-a
11-a
Disk
Cntl
Disk
14-b
11-b
Disk
Cntl
Disk

【図７】

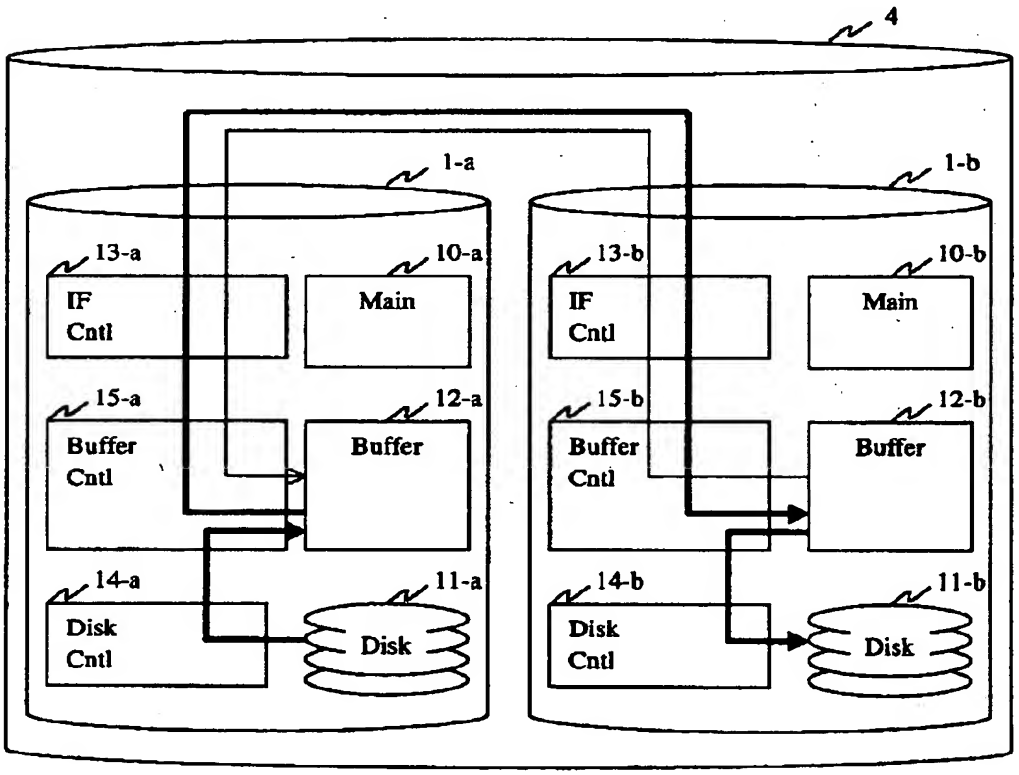
図７

30

#	未更新開始アドレス	未更新サイズ	更新情報
0			
1			
2			
・			
N			

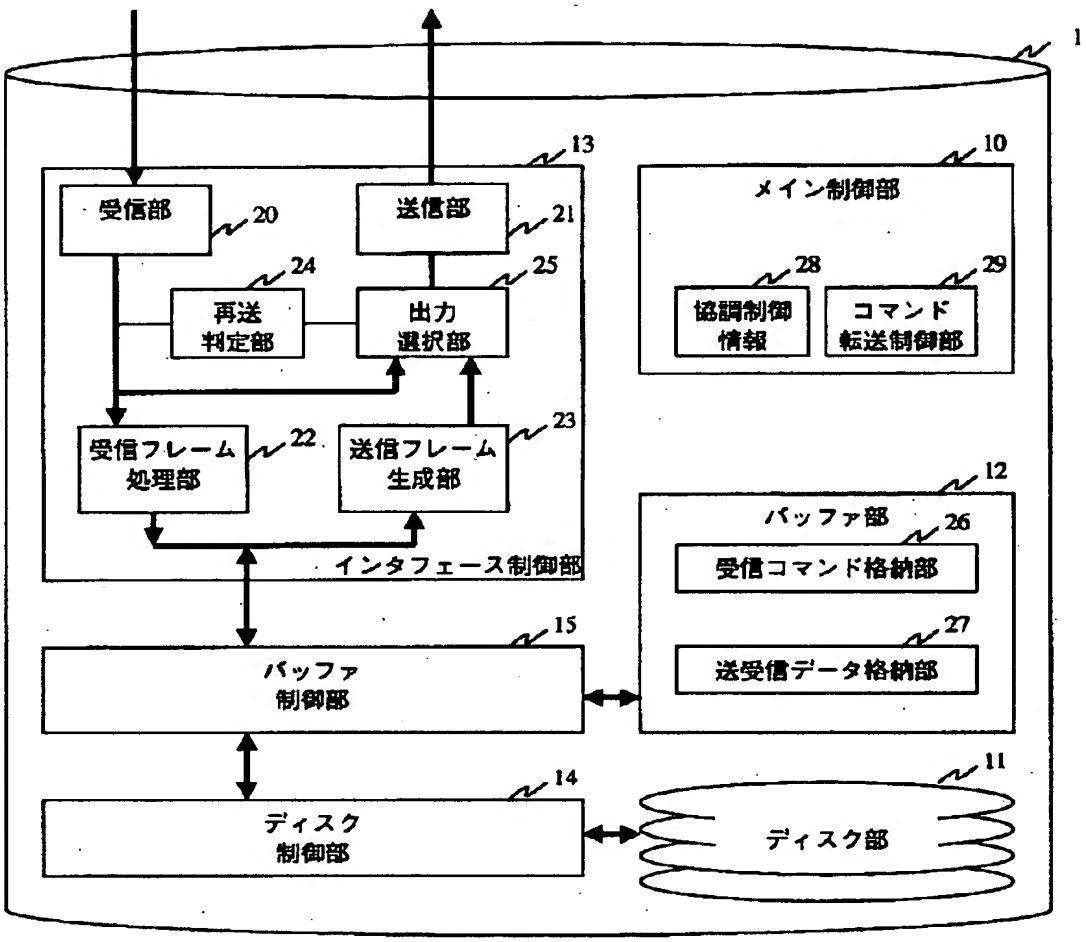
【図８】

図８



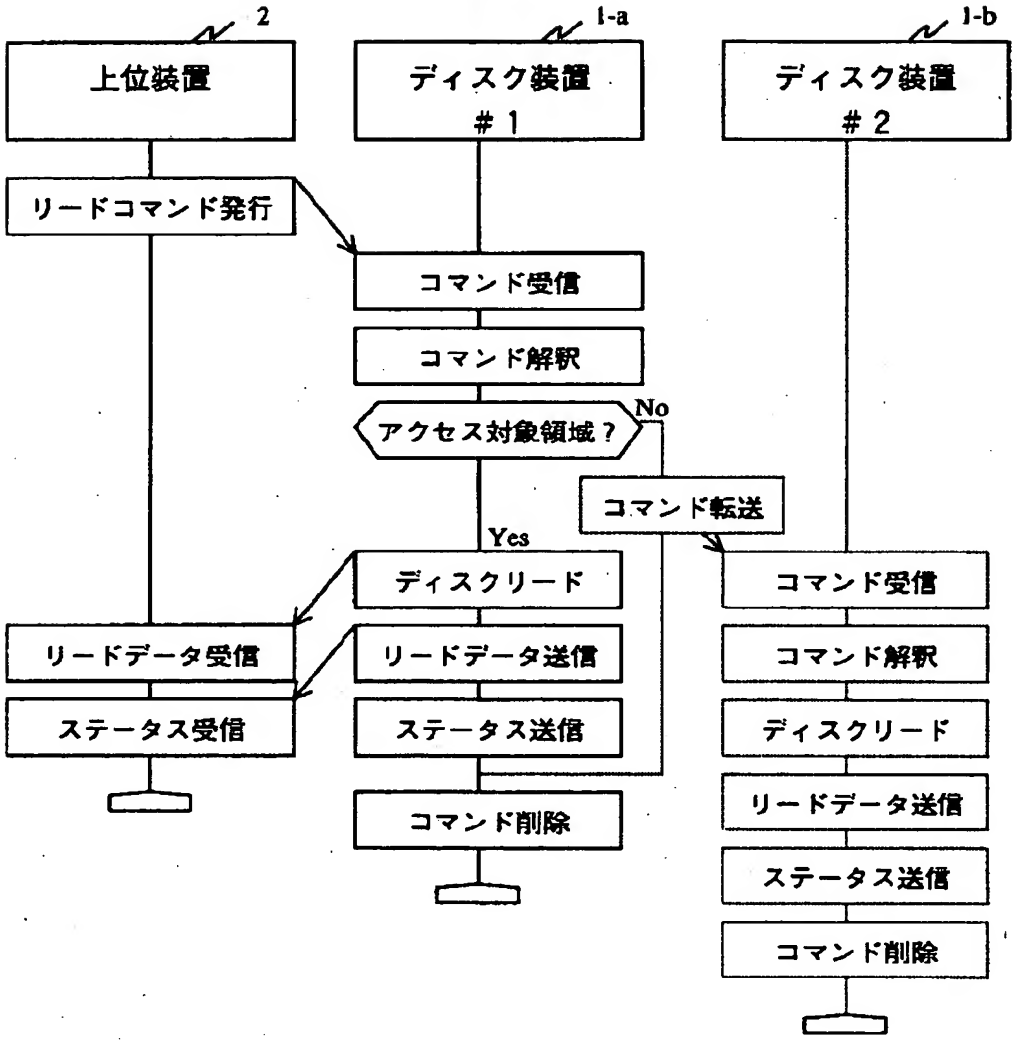
【図９】

図９


1
13
受信部
20
送信部
21
24
再送
判定部
25
出力
選択部
受信フレーム
処理部
22
送信フレーム
生成部
23
インタフェース制御部
10
メイン制御部
28
協調制御
情報
29
コマンド
転送制御部
12
バッファ部
26
受信コマンド格納部
27
送受信データ格納部
15
バッファ
制御部
14
ディスク
制御部
11
ディスク部

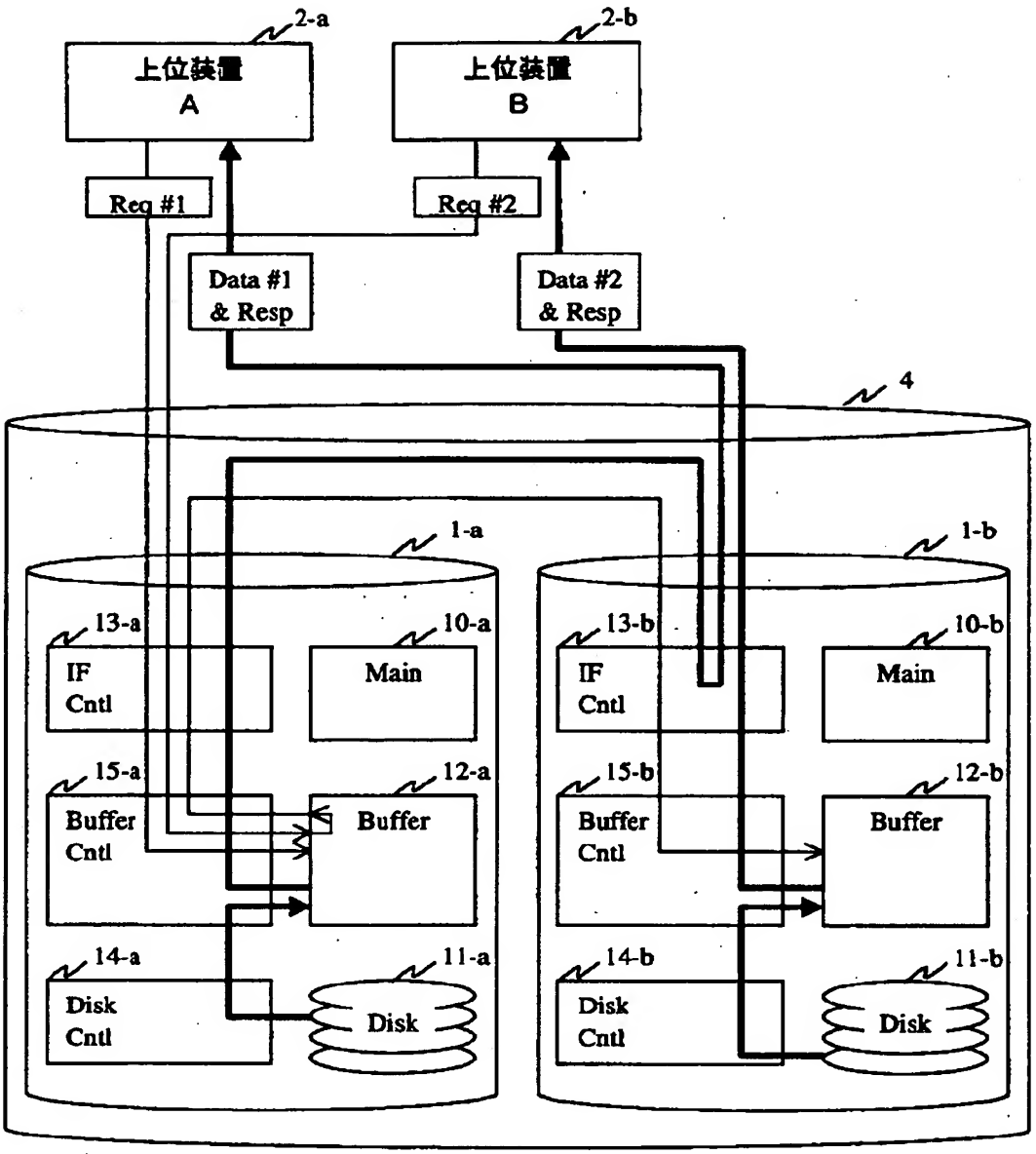
【図１０】

図１０


2
1-a
1-b
上位装置
ディスク装置
＃１
ディスク装置
＃２
リードコマンド発行
コマンド受信
コマンド解釈
アクセス対象領域？
No
Yes
コマンド転送
ディスクリード
コマンド受信
リードデータ受信
リードデータ送信
コマンド解釈
ステータス受信
ステータス送信
ディスクリード
コマンド削除
リードデータ送信
ステータス送信
コマンド削除

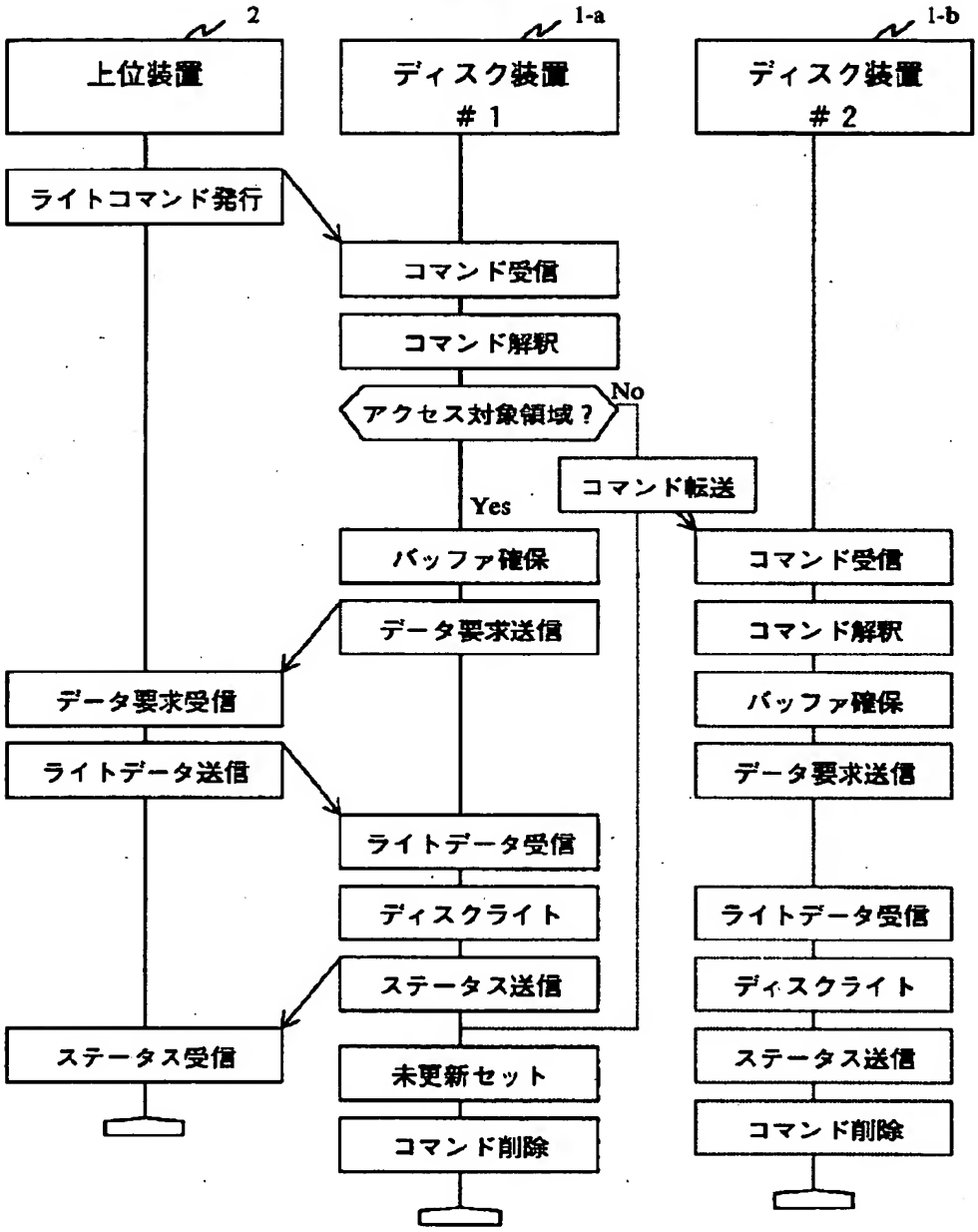
【図１１】

図１１


2-a
上位装置
A
2-b
上位装置
B
Req #1
Req #2
Data #1
& Resp
Data #2
& Resp
4
1-a
1-b
13-a
IF
Cntl
10-a
Main
13-b
IF
Cntl
10-b
Main
15-a
Buffer
Cntl
12-a
Buffer
15-b
Buffer
Cntl
12-b
Buffer
14-a
Disk
Cntl
11-a
Disk
14-b
Disk
Cntl
11-b
Disk

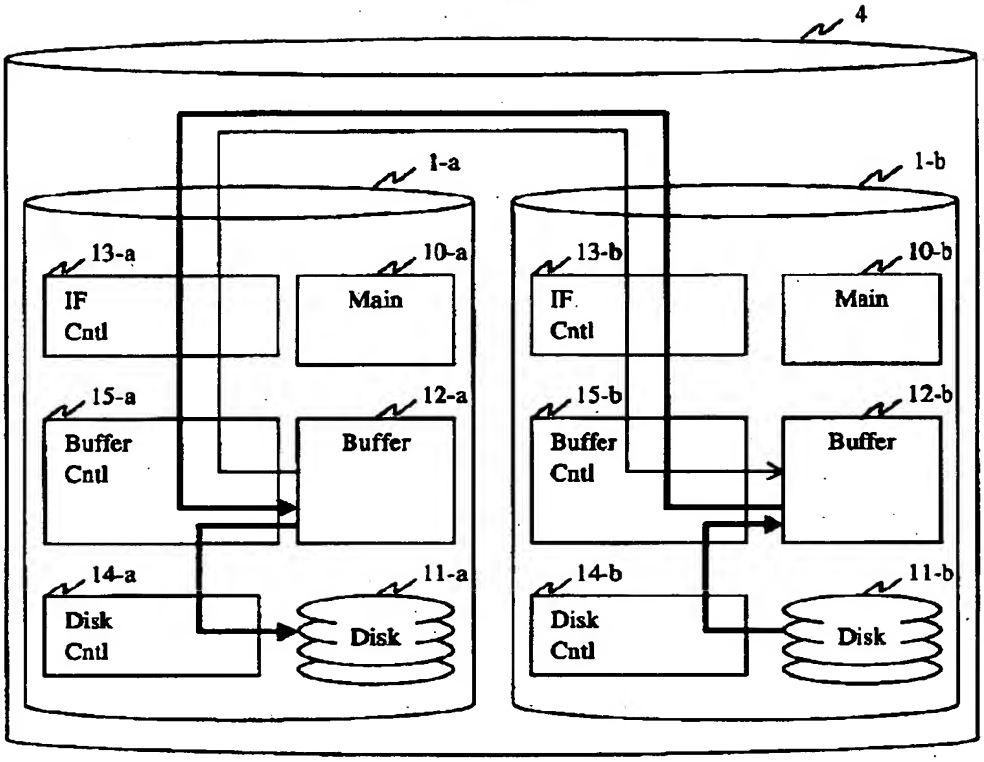
【図１２】

図１２


2
1-a
1-b
上位装置
ディスク装置
＃１
ディスク装置
＃２
ライトコマンド発行
コマンド受信
コマンド解釈
アクセス対象領域？
No
Yes
コマンド転送
バッファ確保
データ要求送信
コマンド受信
コマンド解釈
バッファ確保
データ要求送信
データ要求受信
ライトデータ送信
ライトデータ受信
ディスクライト
ライトデータ受信
ディスクライト
ステータス送信
ステータス送信
ステータス受信
未更新セット
コマンド削除
コマンド削除

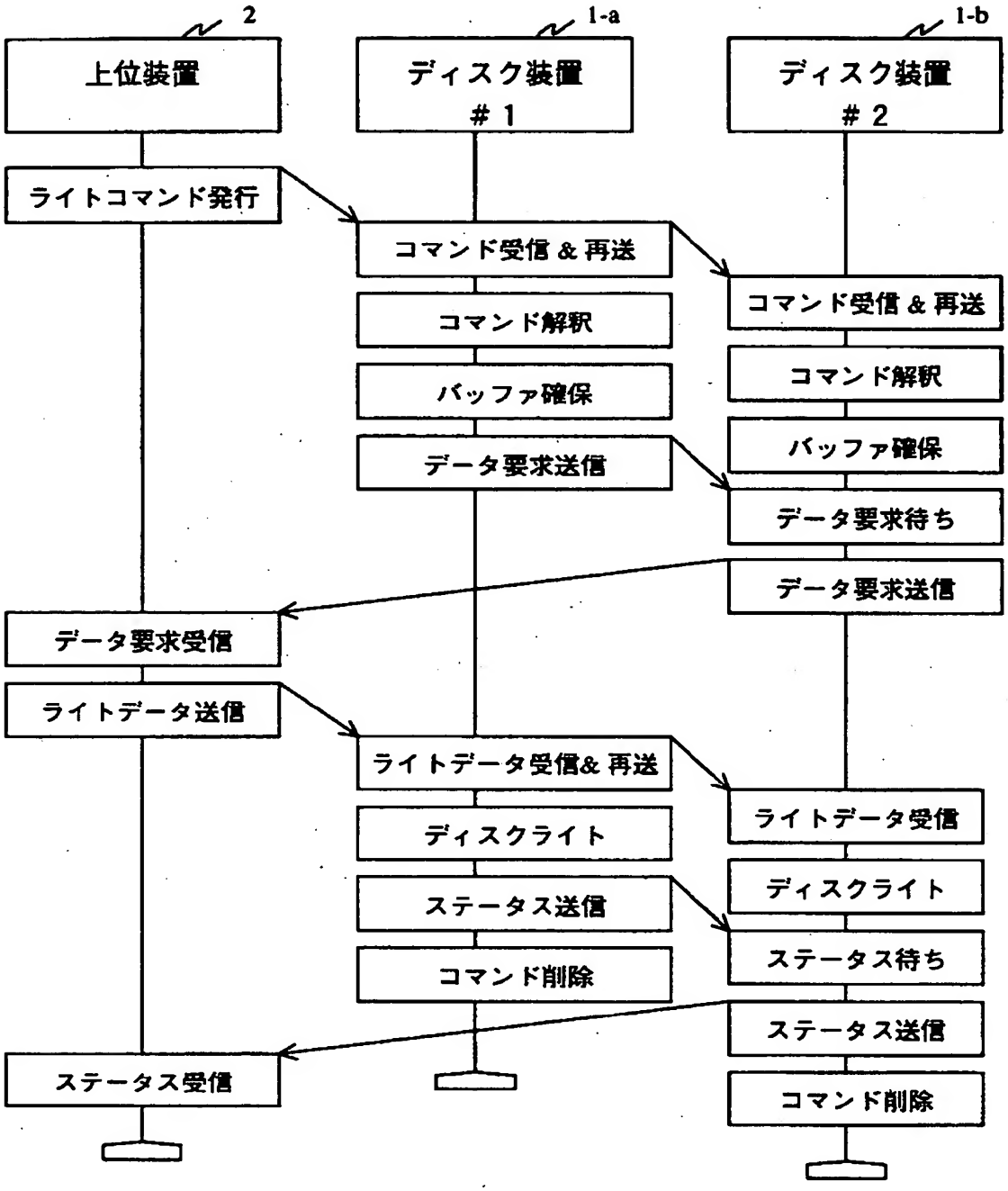
【図１３】

図１３


4
1-a
1-b
13-a
IF
Cntl
10-a
Main
15-a
Buffer
Cntl
12-a
Buffer
14-a
Disk
Cntl
11-a
Disk
13-b
IF
Cntl
10-b
Main
15-b
Buffer
Cntl
12-b
Buffer
14-b
Disk
Cntl
11-b
Disk

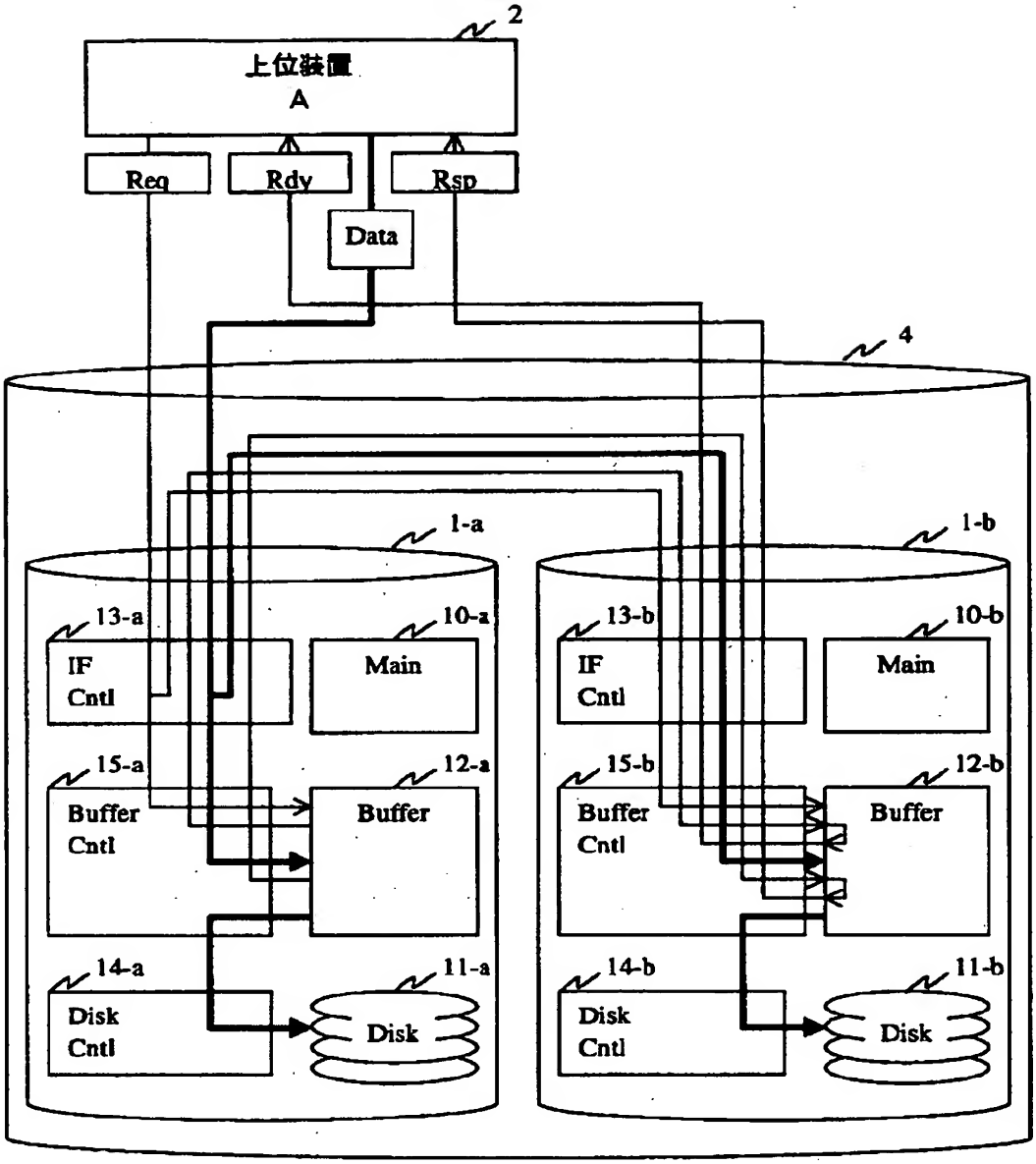
【図１４】

図１４


2
1-a
1-b
上位装置
ディスク装置
＃１
ディスク装置
＃２
ライトコマンド発行
コマンド受信＆再送
コマンド解釈
バッファ確保
データ要求送信
コマンド受信＆再送
コマンド解釈
バッファ確保
データ要求待ち
データ要求送信
データ要求受信
ライトデータ送信
ライトデータ受信&再送
ディスクライト
ステータス送信
コマンド削除
ライトデータ受信
ディスクライト
ステータス待ち
ステータス送信
コマンド削除
ステータス受信

【図１５】

図１５


2
上位装置
A
Req
Rdy
Rsp
Data
4
1-a
1-b
13-a
10-a
IF
Cntl
Main
15-a
12-a
Buffer
Cntl
Buffer
14-a
11-a
Disk
Cntl
Disk
13-b
10-b
IF
Cntl
Main
15-b
12-b
Buffer
Cntl
Buffer
14-b
11-b
Disk
Cntl
Disk

【図１６】

図１６

協調制御情報項目		記憶装置＃１設定例	記憶装置＃２設定例
記憶装置システムID（共有ID）		ID_0	ID_0
記憶装置ID（固有ID）		ID_1	ID_2
協調装置ID（協調記憶装置固有ID）		ID_2	ID_1
協調処理（コマンド転送）モード		コマンド再送	コマンド受領
領域管理情報	協調処理領域	0 - (N-1)	0 - (N-1)
	アクティブ領域開始アドレス	0	N/2
	アクティブ領域サイズ	N/2	N/2

【図１７】

図１７

Bit / Byte	7	6	5	4	3	2	1	0
0 - 7	FCP_LUN							
8	Command Reference Number							
9	Reserved						Task Attribute	
10	Task Management Flags							
11	Reserved						Rd Data	Wr Data
12	Operation Code							
13	Logical Unit Number			Flag				
14 - 17	Logical Block Address							
18	Reserved							
19 - 20	Transfer Length							
21	Control							
22 - 27	Reserved							
28 - 31	FCP_DL							

【図１８】

図１８

Byte / Word	3	2	1	0
0	R_CTL	D_ID		
1	CS_CTL	S_ID		
2	TYPE	F_CTL		
3	SEQ_ID	SEQ_ID	SEQ_CNT	
4	OX_ID		RX_ID	
5	Parameter			

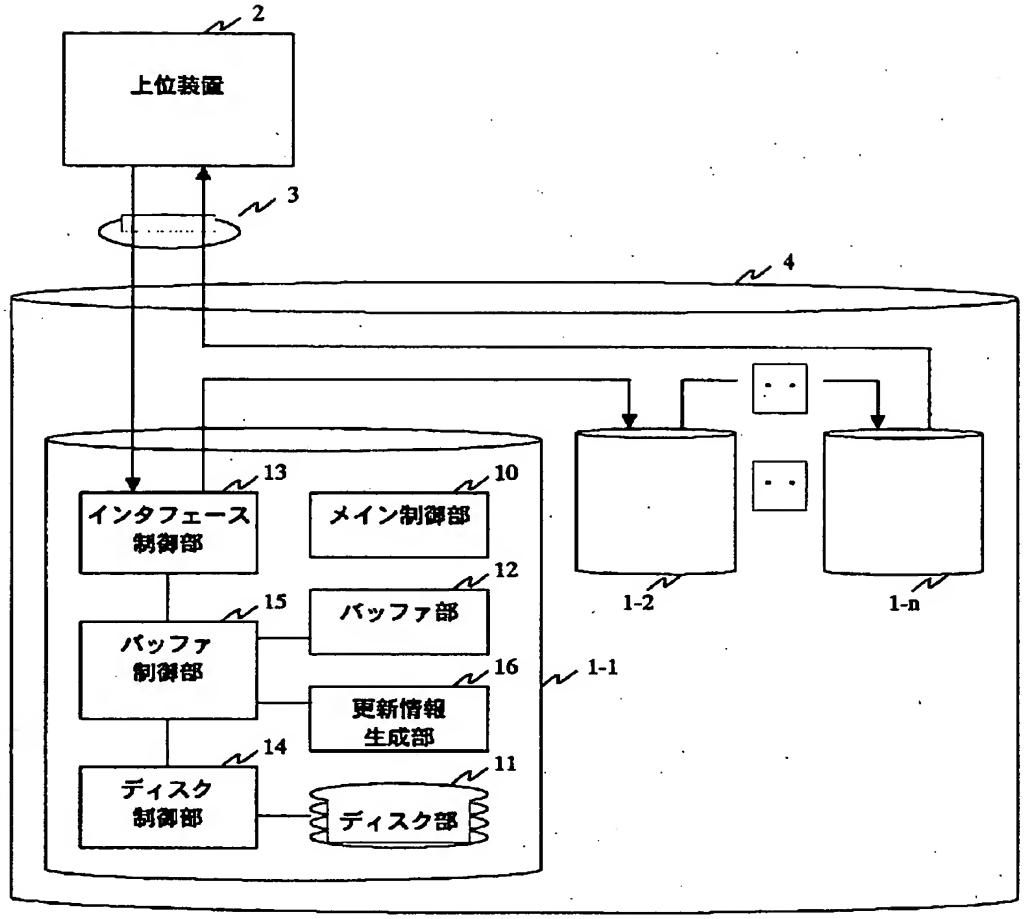
【図１９】

図１９

#	Host ID (S_ID)	OX_ID	RX_ID
0	ID_3	OX_00	RX_00
1	ID_3	OX_01	RX_02
2	ID_4	OX_01	RX_04
-	-	-	-
N	-	-	-

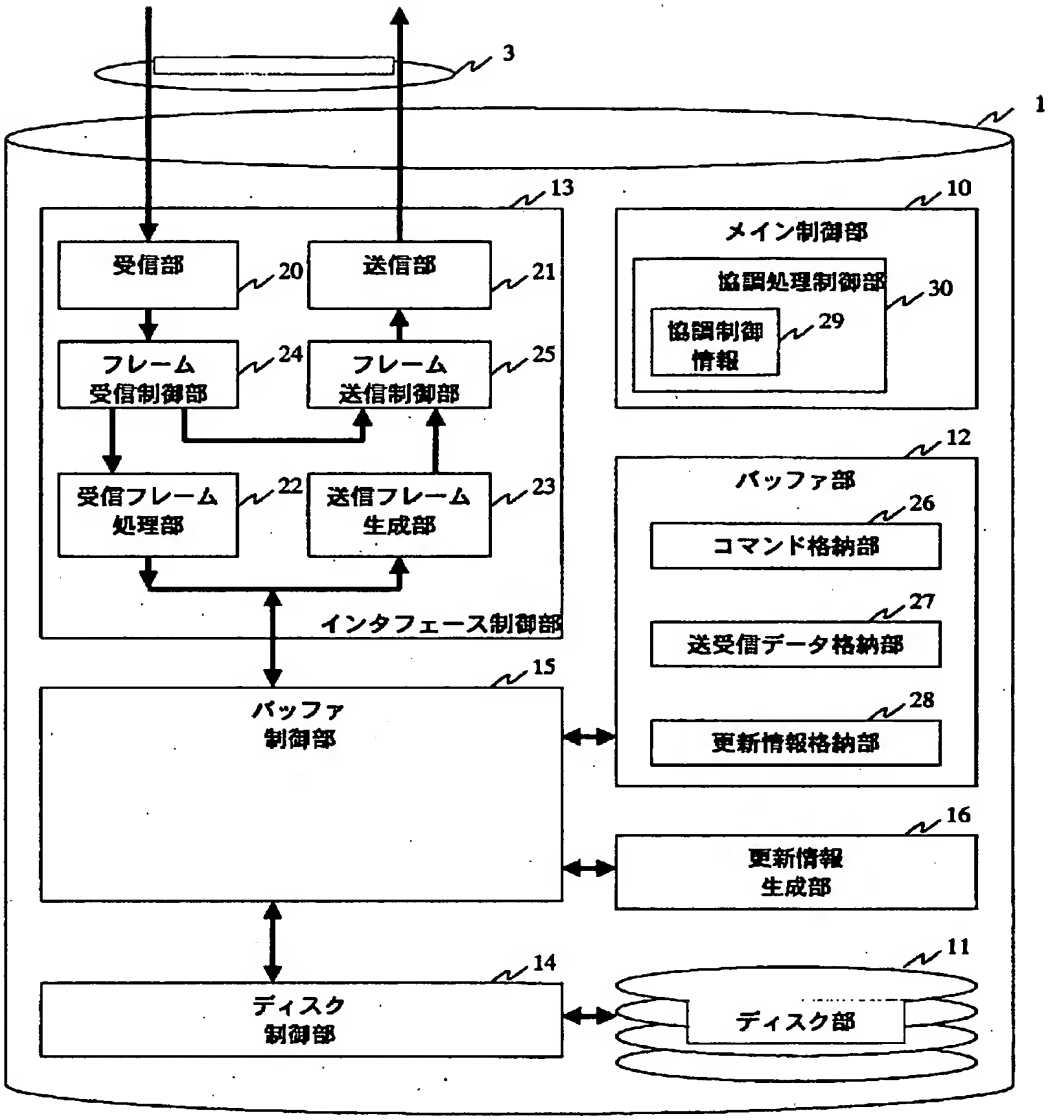
【図２０】

図２０


2
上位装置
3
4
13
インタフェース
制御部
10
メイン制御部
12
バッファ部
15
バッファ
制御部
16
更新情報
生成部
1-1
14
ディスク
制御部
11
ディスク部
1-2
1-n

【図２１】

図２１


3
1
13
10
受信部
20
送信部
21
メイン制御部
協調処理制御部
30
協調制御
情報
29
フレーム
受信制御部
24
フレーム
送信制御部
25
12
バッファ部
受信フレーム
処理部
22
送信フレーム
生成部
23
26
コマンド格納部
27
送受信データ格納部
インタフェース制御部
15
28
更新情報格納部
バッファ
制御部
16
更新情報
生成部
11
14
ディスク
制御部
ディスク部

【図２２】

図２２

協調制御情報項目		記憶装置＃１	記憶装置＃２	記憶装置＃Ｎ
記憶装置サブシステムID（共有ID）		ID_0	←	←
記憶装置ID（固有ID）		ID_1	ID_2	ID_N
ポジションマップ		1	2	N
コマンド転送モード		受信＆再送	受信＆再送	受信のみ
冗長構成管理情報	RAIDレベル	5	←	←
	データディスク台数	Ｎ－１	←	←
	冗長データディスク台数	1	←	←
	冗長データ管理サイズ	６４（kB）	←	←

【図２３】

図２３

SOF	Header	Payload	CRC	EOF

【図２４】

図２４

Byte / Word	3	2	1	0
0	R_CTL	D_ID		
1	CS_CTL	S_ID		
2	TYPE	F_CTL		
3	SEQ_ID	SEQ_ID	SEQ_CNT	
4	OX_ID		RX_ID	
5	Parameter			

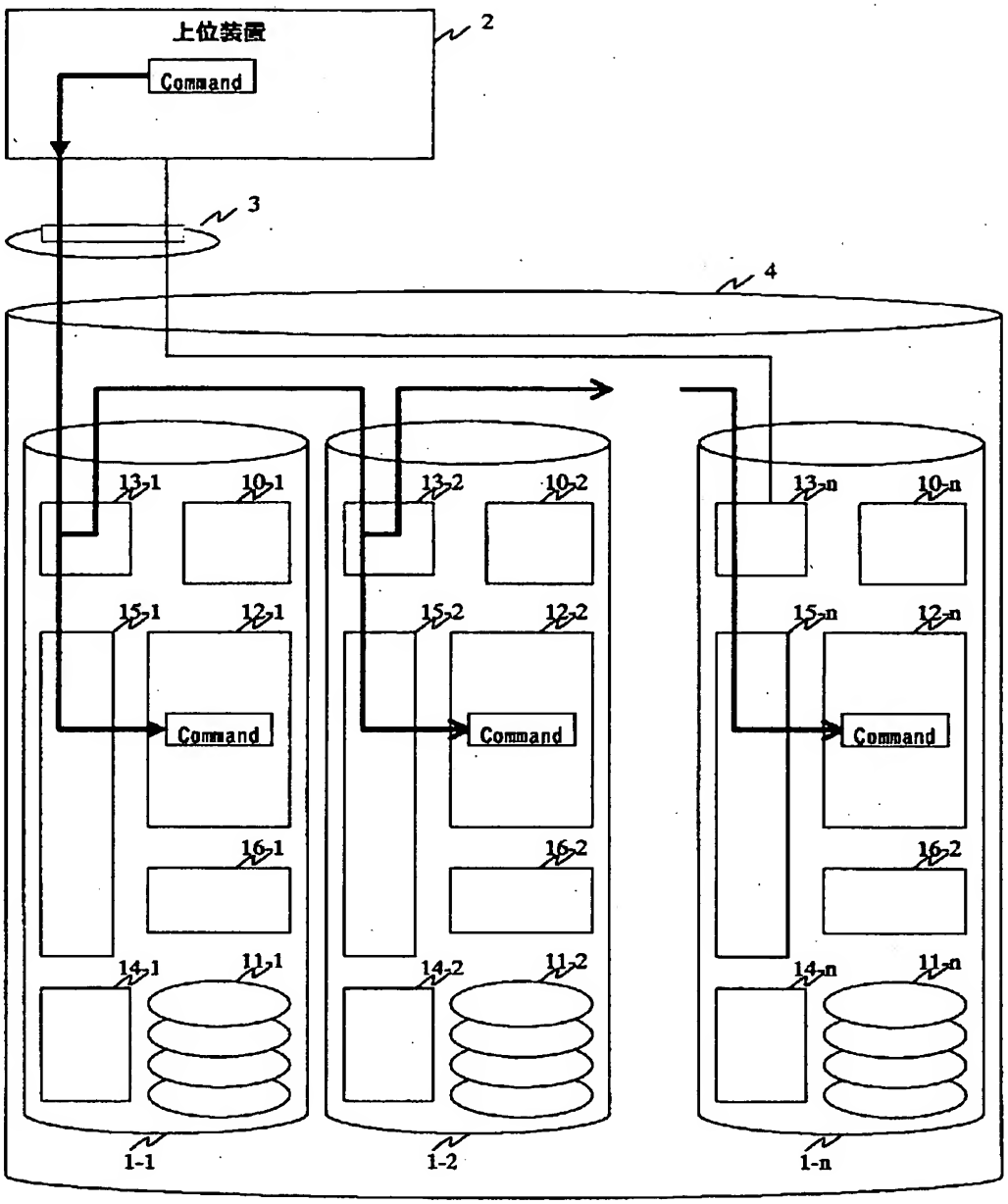
【図２５】

図２５

Bit / Byte	7	6	5	4	3	2	1	0
0 - 7	FCP_LUN							
8	Command Reference Number							
9	Reserved						Task Attribute	
10	Task Management Flags							
11	Reserved						Rd Data	Wr Data
12	Operation Code							
13	Logical Unit Number				Flag			
14 - 17	Logical Block Address							
18	Reserved							
19 - 20	Transfer Length							
21	Control							
22 - 27	Reserved							
28 - 31	FCP_DL							

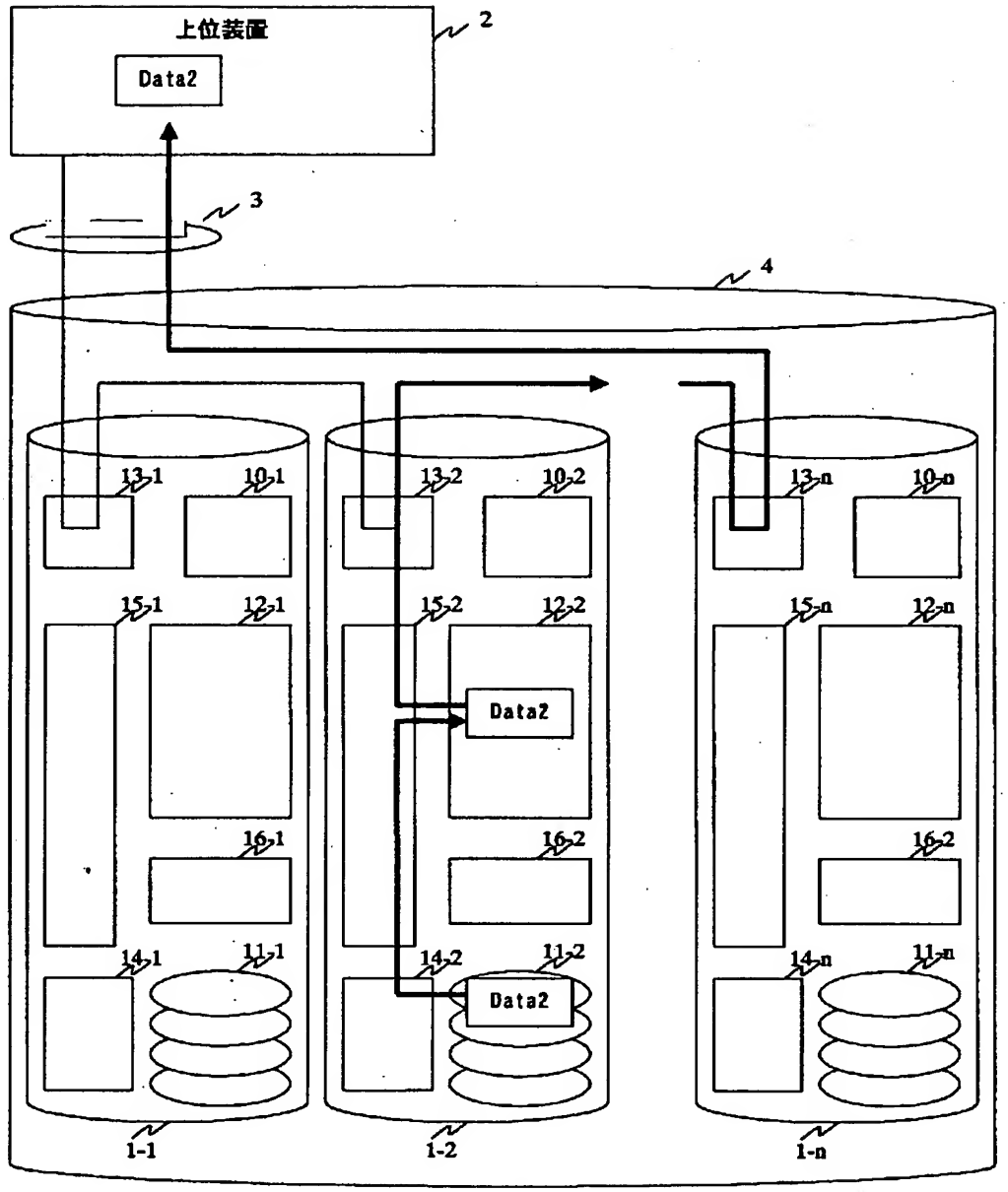
【図２６】

図２６


上位装置
Command
2
3
4
13-1
10-1
13-2
10-2
13-n
10-n
15-1
12-1
15-2
12-2
15-n
12-n
Command
Command
Command
16-1
16-2
16-2
14-1
11-1
14-2
11-2
14-n
11-n
1-1
1-2
1-n

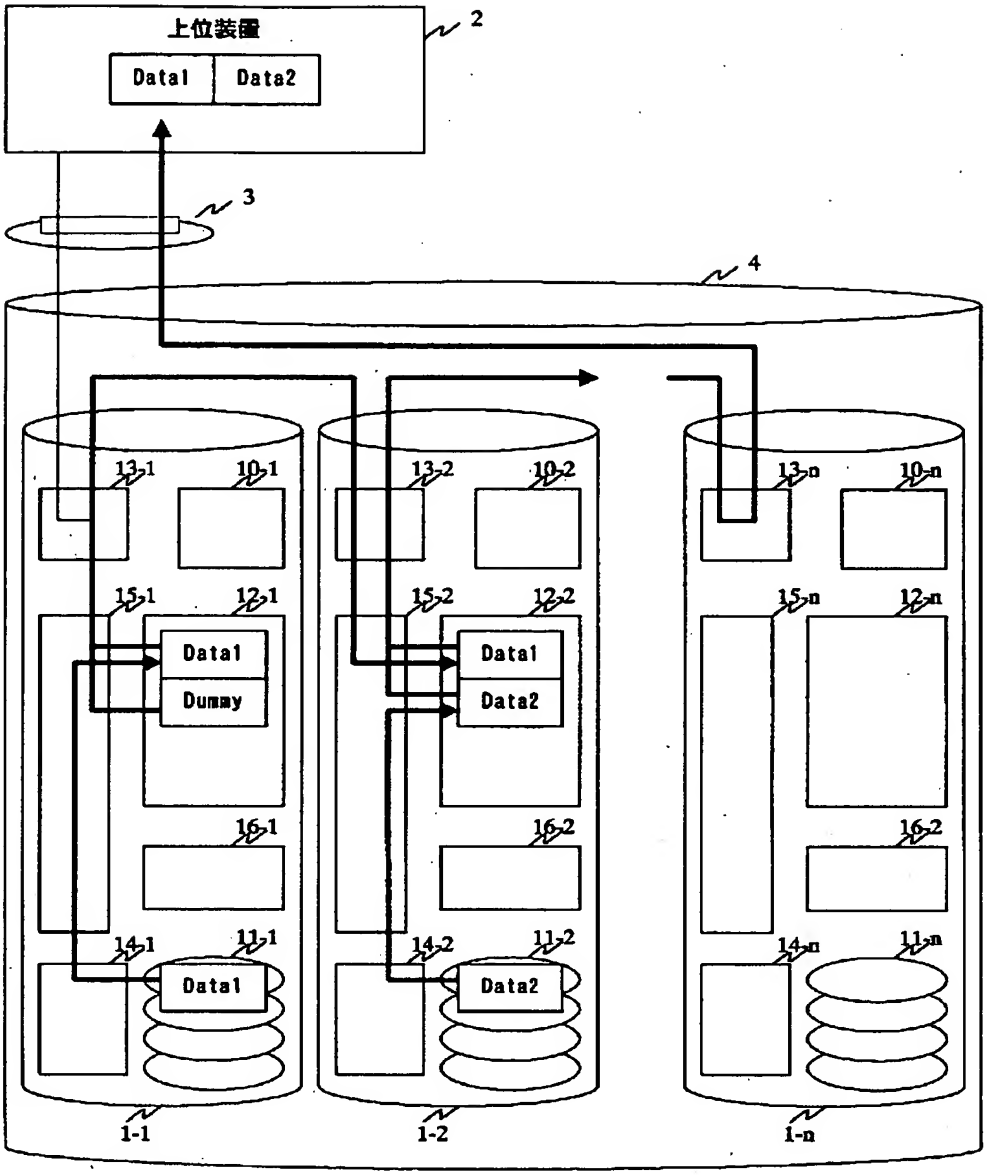
【図２７】

図２７


上位装置
Data2
2
3
4
13-1
10-1
13-2
10-2
13-n
10-n
15-1
12-1
15-2
12-2
15-n
12-n
Data2
16-1
16-2
16-2
14-1
11-1
14-2
11-2
14-n
11-n
Data2
1-1
1-2
1-n

【図２８】

図２８


上位装置
Data1
Data2
2
3
4
13-1
10-1
13-2
10-2
13-n
10-n
15-1
12-1
15-2
12-2
15-n
12-n
Data1
Dummy
Data1
Data2
16-1
16-2
16-2
14-1
11-1
14-2
11-2
14-n
11-n
Data1
Data2
1-1
1-2
1-n

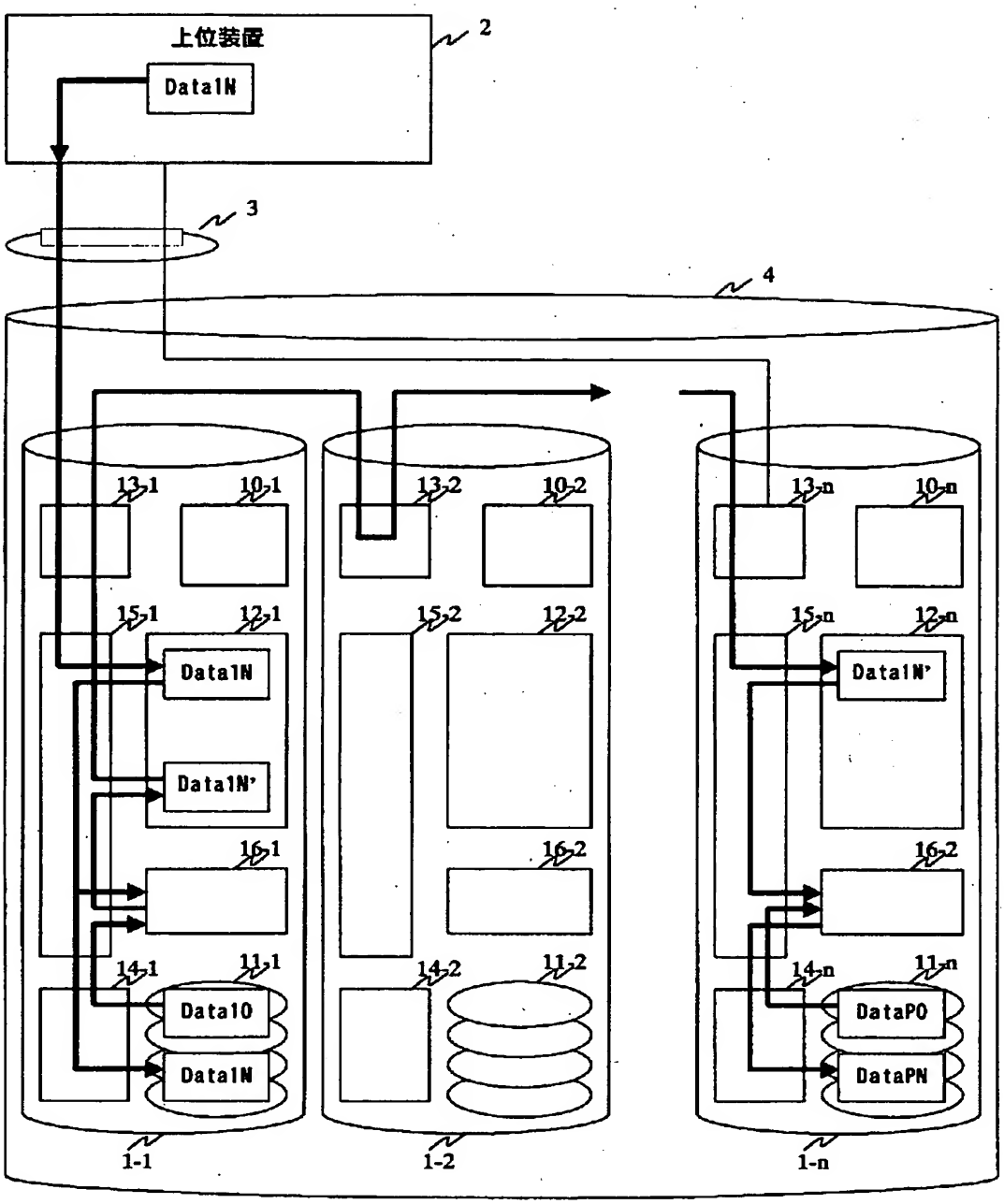
【図２９】

図２９

Bit / Byte	7	6	5	4	3	2	1	0
0 - x	Data Valid bit							
-	User (Read/Write) Data Segment 0							
-	:							
-	User (Read/Write) Data Segment n							

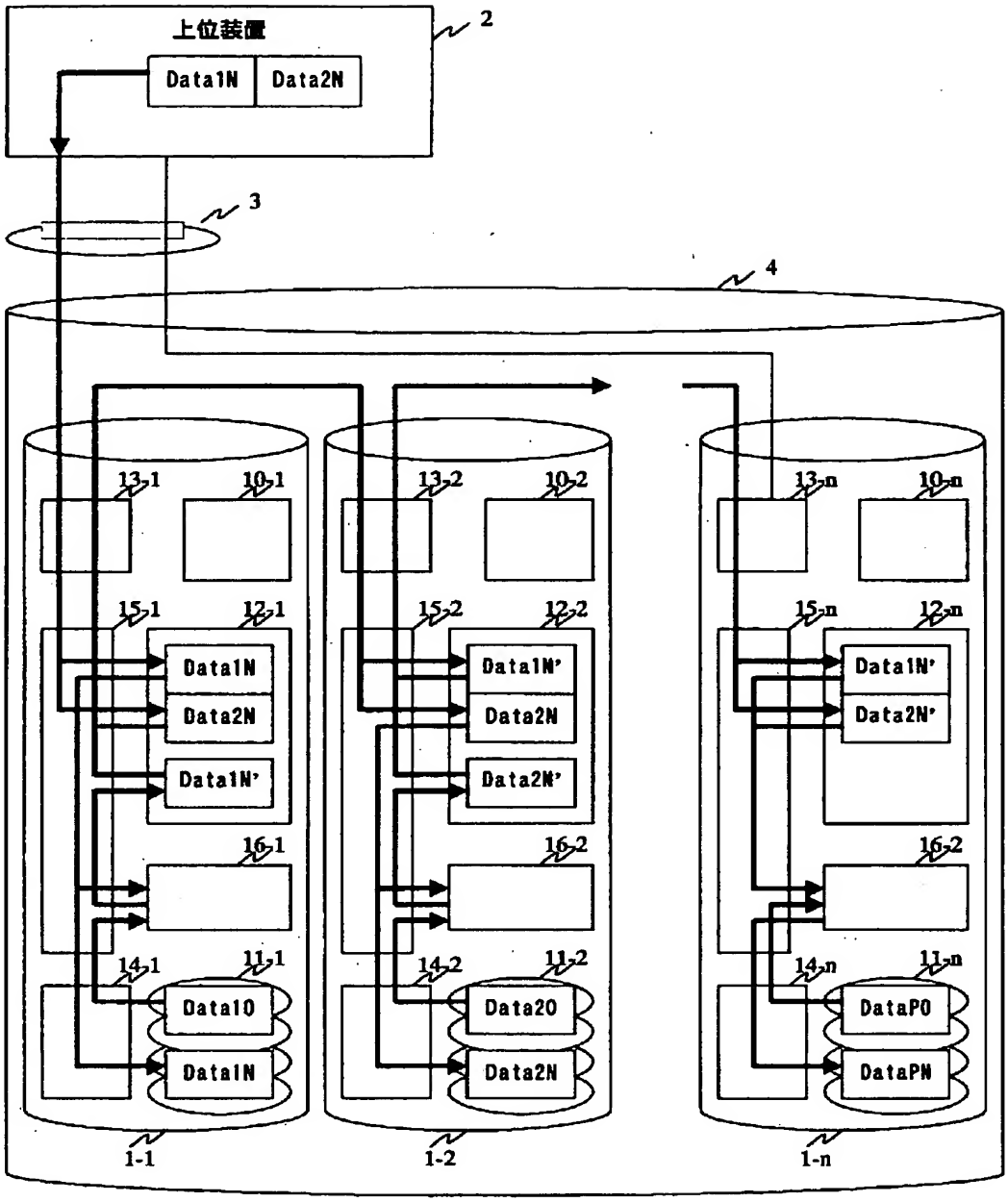
【図３０】

図３０


上位装置
2
DataIN
3
4
13-1
10-1
13-2
10-2
13-n
10-n
15-1
12-1
15-2
12-2
15-n
12-n
DataIN
DataIN'
DataIN'
16-1
16-2
16-2
14-1
11-1
14-2
11-2
14-n
11-n
DataIO
DataIN
DataPO
DataPN
1-1
1-2
1-n

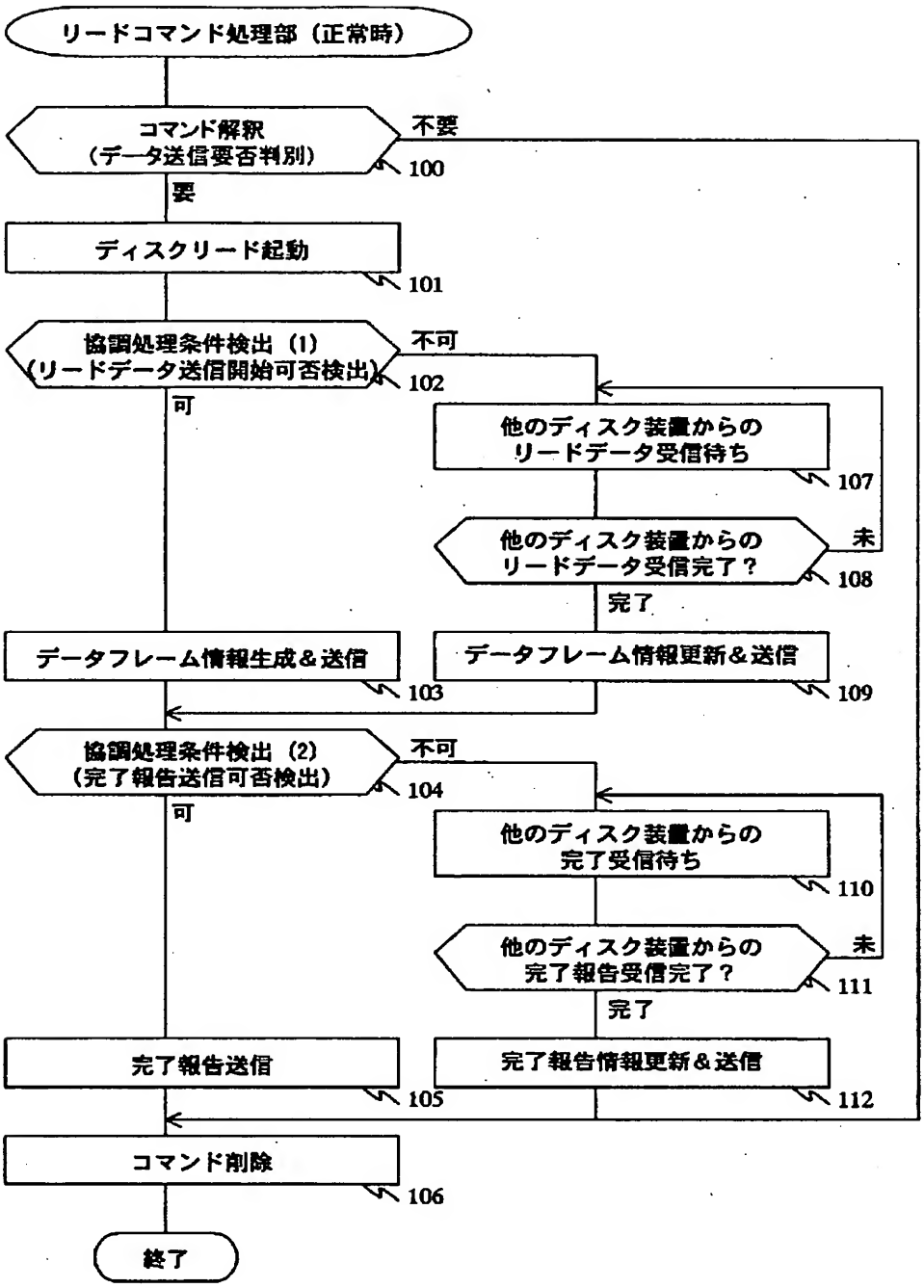
【図３１】

図３１


上位装置
Data1N
Data2N
2
3
4
13-1
10-1
13-2
10-2
13-n
10-n
15-1
12-1
15-2
12-2
15-n
12-n
Data1N
Data2N
Data1N'
Data1N'
Data2N
Data2N'
Data1N'
Data2N'
16-1
16-2
16-2
14-1
11-1
14-2
11-2
14-n
11-n
Data10
Data1N
Data20
Data2N
DataP0
DataPN
1-1
1-2
1-n

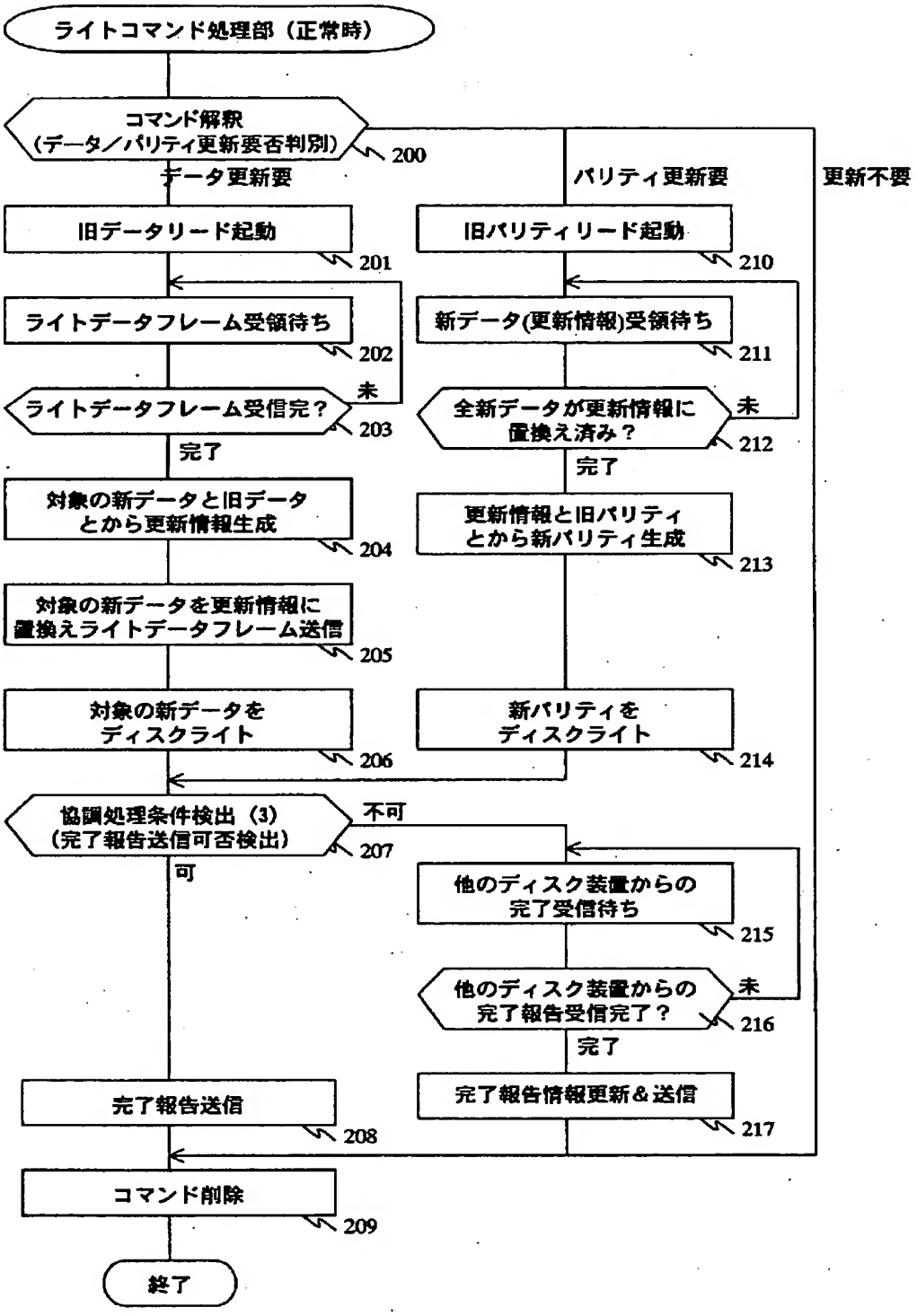
【図３２】

図３２


リードコマンド処理部（正常時）
コマンド解釈
（データ送信要否判別）
100
不要
要
ディスクリード起動
101
協調処理条件検出（1）
（リードデータ送信開始可否検出）
102
不可
可
他のディスク装置からの
リードデータ受信待ち
107
他のディスク装置からの
リードデータ受信完了？
108
未
完了
データフレーム情報生成＆送信
103
データフレーム情報更新＆送信
109
協調処理条件検出（2）
（完了報告送信可否検出）
104
不可
可
他のディスク装置からの
完了受信待ち
110
他のディスク装置からの
完了報告受信完了？
111
未
完了
完了報告送信
105
完了報告情報更新＆送信
112
コマンド削除
106
終了

【図３３】

図３３


ライトコマンド処理部（正常時）
コマンド解釈
（データ／パリティ更新要否判別）
200
データ更新要
パリティ更新要
更新不要
旧データリード起動
201
旧パリティリード起動
210
ライトデータフレーム受領待ち
202
新データ(更新情報)受領待ち
211
ライトデータフレーム受信完？
203
未
全新データが更新情報に
置換え済み？
212
未
完了
完了
対象の新データと旧データ
とから更新情報生成
204
更新情報と旧パリティ
とから新パリティ生成
213
対象の新データを更新情報に
置換えライトデータフレーム送信
205
対象の新データを
ディスクライト
206
新パリティを
ディスクライト
214
協調処理条件検出（3）
（完了報告送信可否検出）
207
不可
可
他のディスク装置からの
完了受信待ち
215
他のディスク装置からの
完了報告受信完了？
216
未
完了
完了報告送信
208
完了報告情報更新＆送信
217
コマンド削除
209
終了

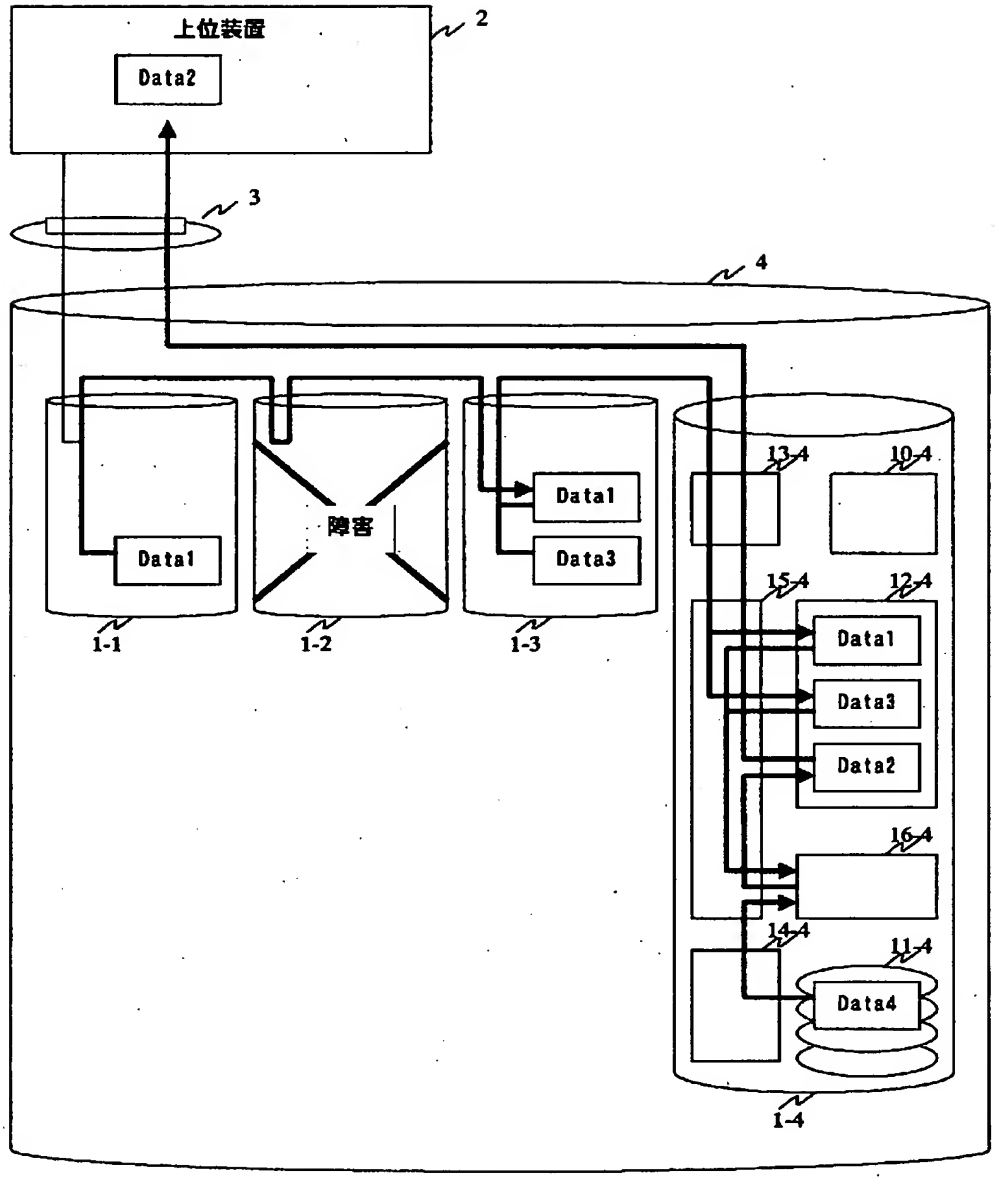
【図３４】

図３４

協調制御情報項目		記憶装置＃１	記憶装置＃２	記憶装置＃Ｎ
記憶装置サブシステムID（共有ID）		ID_0	←	←
記憶装置ID（固有ID）		ID_1	ID_2	ID_N
ポジションマップ		1	2	N
コマンド転送モード		受信＆再送	受信＆再送	受信のみ
冗長構成管理情報	RAIDレベル	5	←	←
	データディスク台数	3	←	←
	冗長データディスク台数	1	←	←
	冗長データ管理サイズ	６４（kB）	←	←
	動作モード（正常／縮退）	縮退	←	←
	障害記憶装置ID	ID_2	←	←

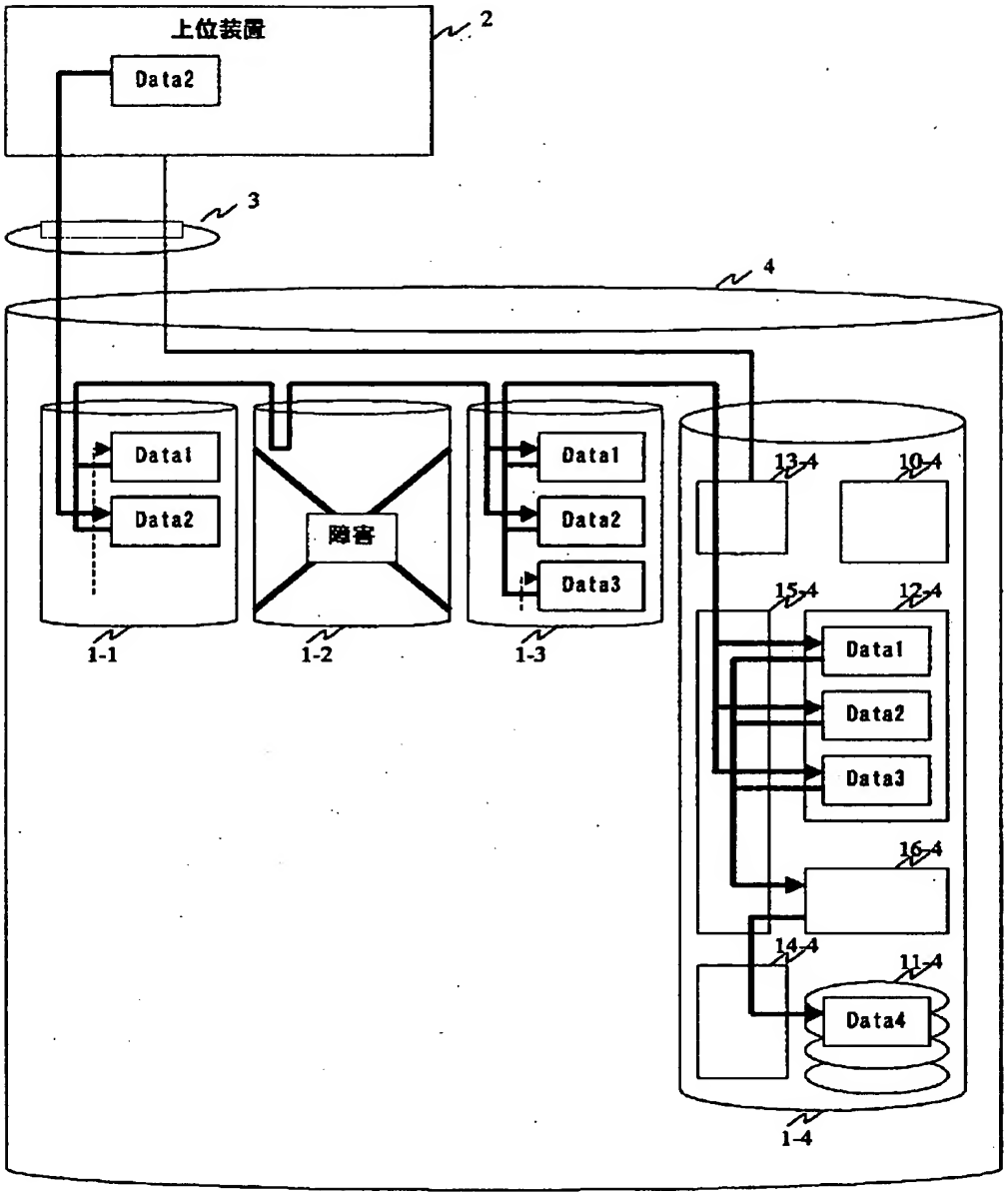
【図３５】

図３５


上位装置
Data2
2
3
4
Data1
障害
Data1
Data3
1-1
1-2
1-3
13-4
10-4
15-4
12-4
Data1
Data3
Data2
16-4
14-4
11-4
Data4
1-4

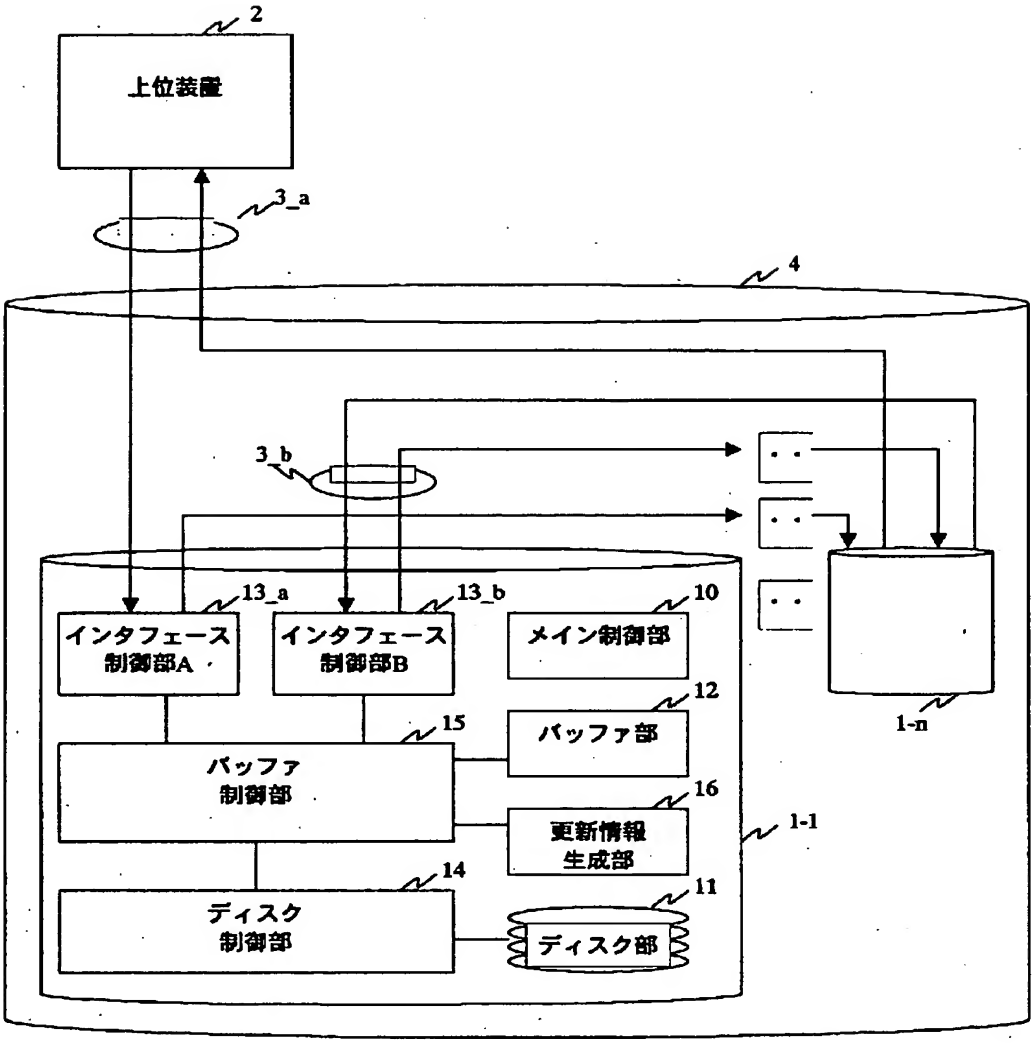
【図３６】

図３６


上位装置
2
Data2
3
4
Data1
Data2
障害
Data1
Data2
Data3
1-1
1-2
1-3
13-4
10-4
15-4
12-4
Data1
Data2
Data3
16-4
14-4
11-4
Data4
1-4

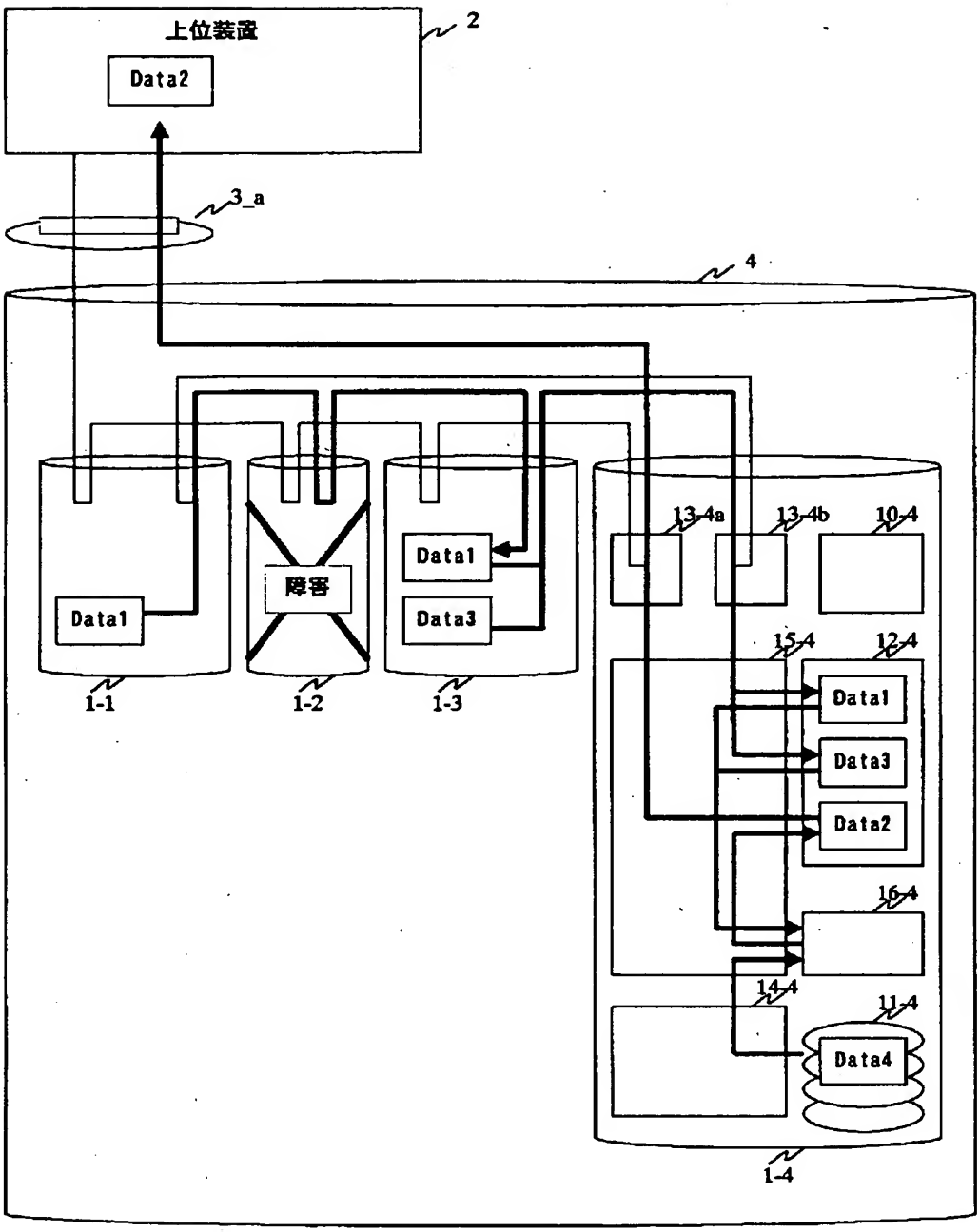
【図３７】

図３７


2
上位装置
3_a
4
3_b
13_a
13_b
10
インタフェース
制御部A
インタフェース
制御部B
メイン制御部
12
バッファ部
15
バッファ
制御部
16
更新情報
生成部
1-1
14
11
ディスク
制御部
ディスク部
1-n

【図３８】

図３８


上位装置
Data2
2
3_a
4
Data1
障害
Data1
Data3
1-1
1-2
1-3
13-4a
13-4b
10-4
15-4
12-4
Data1
Data3
Data2
16-4
14-4
11-4
Data4
1-4

【書類名】　要約書

【要約】

【課題】

複数の記憶装置から構成されるRAID1の記憶装置システムにおいて、上位装置からのアクセス要求を、該記憶装置システムを構成する複数の記憶装置間で協調して処理する制御方式を提供する。

【解決手段】

RAID1の記憶装置システムを構成する個々の記憶装置が、上位装置からのアクセス要求を共有する手段と、上記共有したアクセス要求について、自身が処理すべきか否かを判別する手段と、更に、上位装置から記憶装置システムに送信されるライトデータを、少なくとも、処理すべき記憶装置に転送する手段とを具備する。

【選択図】　図２

認定・付加情報

特許出願の番号	特願２００１－３４４０１０
受付番号	５０１０１６５４６６９
書類名	特許願
担当官	第七担当上席　　００９６
作成日	平成１３年１１月１４日

＜認定情報・付加情報＞

【提出日】　平成13年11月 9日

次頁無

出　願　人　履　歴　情　報

識別番号　　〔０００００５１０８〕

１．変更年月日　　１９９０年　８月３１日
［変更理由］　新規登録
住　所　　東京都千代田区神田駿河台４丁目６番地
氏　名　　株式会社日立製作所